Proceedings

ICSIIT 2012

International Conference on
Soft Computing, Intelligent System and Information Technology

24-25 May 2012

Bali, Indonesia

Editors:
Leo Willyanto Santoso
Andreas Handojo

Informatics Department
Petra Christian University

Center of Soft Computing and
Intelligent System Studies

Proceedings

International Conference on Soft Computing, Intelligent System and Information Technology 2012

Publisher:
Informatics Department
Petra Christian University

ISBN: 978-602-97124-1-4

Additional copies may be ordered from:
Informatics Department
Petra Christian University, Siwalankerto 121-131, Surabaya 60236, Indonesia

Cover Art production by Adi Wibowo/Informatics Department

ICSIIT 2012
Table of Contents

Imaging & Multimedia Technology

Computer Network, Mobile Application, Web Services & Security

Application of Information System & Technology

Preface

First of all, I would like to give thank to God the Creator, God the Redeemer and God who leads us to the truth for all His blessings to us. As we all know, this 3rd International Conference on Soft Computing, Intelligent Systems and Information Technology 2012 (ICSIIT 2012) is held from 24-25 May 2012 in the Inna Kuta Beach Hotel located at this paradise island, Bali, Indonesia. I thank Him for His presence and guidance in letting this conference happen. Only by God's grace, we hope we could give our best for 3rd ICSIIT 2012 despite of all of our limitation.

We thank all authors who have contributed and participated in presenting their works at this conference. We also gratefully acknowledge the important review supports provided by the 16 members of the program committee from 10 different countries. Their efforts were crucial to the success of the conference. We are also so blessed by the presence of keynote speaker who will address the important trends relating medical imaging and soft computing. Prof. Sankar Kumar Pal, Ph.D. will present "Rough-fuzzy Computation, Pattern Recognition and Data Mining: Application to medical imaging and bioinformatics".

I hope during your stay in this beautiful island you will enjoy and benefit both, the fresh sea breeze and harmonious sound from sea waves, as well as the intellectual and scientific discussions. I hope your contributions and participation of the discussion will lead to the benefit of the advancements on Soft Computing, Intelligent Systems and Information Technology.

Soli Deo Gloria,
Adi Wibowo
Conference Chair
ICSIIT 2012 Bali Indonesia

Organizing Committee

The first ICSIIT 2012 is organized by Informatics Department, in cooperation with the Center of Soft Computing and Intelligent System Studies, Petra Christian University, Indonesia.

Conference Chair:

Adi Wibowo	Petra Christian University, Indonesia
Gregorius Satia Budhi	Petra Christian University, Indonesia

Organizing Committee:

Agustinus Noertjahjana	Petra Christian University, Indonesia
Alexander Setiawan	Petra Christian University, Indonesia
Andreas Handojo	Petra Christian University, Indonesia
Arlinah Imam Rahardjo	Petra Christian University, Indonesia
Cherry Galatia Ballangan	Petra Christian University, Indonesia
Djoni Haryadi Setiabudi	Petra Christian University, Indonesia
Ibnu Gunawan	Petra Christian University, Indonesia
Justinus Andjarwirawan	Petra Christian University, Indonesia
Kartika Gunadi	Petra Christian University, Indonesia
Leo Willyanto Santoso	Petra Christian University, Indonesia
Liliana	Petra Christian University, Indonesia
Lily Puspa Dewi	Petra Christian University, Indonesia
Silvia Rostianingsih	Petra Christian University, Indonesia
Yulia	Petra Christian University, Indonesia

Program Committee

A.V.Senthil Kumar (India)

Aniati Murni Arymurthy (Indonesia)

Bhakti Satyabudhi (United Kingdom)

Ben Yip (Australia)

Budi Bambang (Indonesia)

Kelvin Cheng (Australia)

Moeljono Widjaja (Indonesia)

M. Rahmat Widyanto (Indonesia)

Pitoyo Hartono (Japan)

Noboru Takagi (Japan)

Rolly Intan (Indonesia)

Rudy Setiono (Singapore)

Taweesak Kijkanjanarat (Thailand)

Willy Susilo (Australia)

Yung-Chen Hung (Taiwan)

Zuwairie Ibrahim (Malaysia)

Rough-fuzzy Computation, Pattern Recognition and Data Mining: Application to medical imaging and bioinformatics

Sankar K. Pal
Center for Soft Computing Research
Indian Statistical Institute
Kolkata 700108, India

ABSTRACT

Different components of machine intelligence are explained. The role of rough sets in uncertainty handling and granular computing is described. The relevance of its integration with fuzzy sets, namely, rough-fuzzy computing, as a stronger paradigm for uncertainty handling, is explained. Different applications of rough granules, significance of f-granulation and other important issues in their implementations are stated. Generalized rough sets using the concept of fuzziness in granules and sets are defined both for equivalence and tolerance relations. These are followed by definitions of different rough-fuzzy entropy measures. Different tasks such as case generation, class-dependent rough-fuzzy granulation for classification, rough-fuzzy clustering, and measuring image ambiguity measures for miningare then addressed, explaining the nature and characteristics of granules used therein.

While the method of case generation with variable reduced dimension is useful for mining data sets with large dimension and size, class dependent granulation coupled with neighbourhood rough sets for feature selection is efficient in modelling overlapping classes. Significance of a new measure, called "dispersion" of classification performance, which focuses on confused classes for higher level analysis, is explained in this regard. Superiority of rough-fuzzy clustering is illustrated for brain MRI segmentation problem as well as for determining bio-bases in encoding protein sequence for analysis. The former uses c-means whereas the latter is based on c-medoids. Image ambiguity measures take into account the fuzziness in boundary regions, as well as the rough resemblance among nearby gray levels and nearby pixels, and are useful for various image analysis operations. Merits of generalization in rough sets, as well as the incorporation of the concept of rough granulation on the top of fuzziness in gray level are demonstrated for image segmentation problem.

The talk concludes with stating the future directions of research and challenges with other applications including natural computing.

A Fuzzy Time Series Model Based on Neural Networks and Cumulative Probability Distribution Method

Jing-Rong Chang
Department of Information Management Chaoyang University of Technology
168, Jifong East Road, Wufong Township
Taichung 41349, Taiwan (R.O.C.)
chrischang@cyut.edu.tw

Yu-Jie Huang
Department of Information Management Chaoyang University of Technology
168, Jifong East Road, Wufong Township
Taichung 41349, Taiwan (R.O.C.)
s9914612@cyut.edu.tw

ABSTRACT
People usually develop different methods to predict uncertainty problems. In recent years, many studies have proposed different fuzzy time series model to solve forecast problems. In addition, the neural network very popular in the non-linear data modeling, and the binary model is considered beyond the single-variable model [11]. In this paper, we propose a multivariate fuzzy time series method to forecast the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) based on cumulative probability distribution approach (CPDA) and feed forward neural networks to improve forecast accuracy. First, we use the CPDA to construct an appropriate length of the interval in the universe of discourse. Then, the feed forward back-propagation neural networks (BPN) are adopted to train the rules and finally make forecasts for the closing prices of TAIEX. Experimental results show that the predictions of proposed model having better forecasting accuracies than some previous methods.

Keywords
Fuzzy time series, Cumulative probability distribution approach (CPDA), Feed forward back-propagation neural networks (BPN), TAIEX

1. INTRODUCTION
Since Song and Chissom [14][15] proposed the theory of fuzzy time series to solve the restrictions of the traditional time series methods, many studies have proposed different model. In recent years, many methods based on fuzzy time series have been proposed for the forecasting problems, such as enrollment, temperature, and stock index. These existing fuzzy forecasting methods usually partitioned the universe of discourse into several intervals with equal length and utilized the discrete fuzzy number to fuzzification historical data. However, these is an issue of these methods is that do not consider information-intensive to fuzzification historical data, so it's can't effectively reflect the characteristics of the data, and usually affects the forecast performance.

In order to solve above problems, this study propose a fuzzy-neural networks model to solve time series forecasting problem. It adopts cumulative probability distribution approach (CPDA) and feed forward back-propagation neural networks (BPN) to fuzzification data and training the fuzzy rules, and finally makes forecasts. The Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and its corresponding index futures, the Taiwan Futures Exchange (TAIFEX) and stock of trading Volume in Taiwan (VOL) are firstly fuzzified by CPDA. Then, in neural networks process, the fuzzifed data are adopted as inputs to forecast the closing price of daily TAIEX, the results show that our model surpasses in accuracy the models advanced by other models [10][18].

2. PRELIMINARIES
In this section, we briefly introduce fuzzy time series, feed forward neural networks and cumulative probability distribution.

2.1 Fuzzy time series
The concept of fuzzy time series by Song and Chissom proposed [14][15][16]. They first proposed a forecasting model based on Zadeh's works [19] called Fuzzy Times Series, which provided a theoretic framework to model a special dynamic process whose observations are linguistic values. In the following, some basic concepts of fuzzy time series are briefly reviewed [14].

Definition 1: $Y(t)$ (t = …, 0, 1, 2,…) is a subset of a real number. Let $Y(t)$ be the universe of discourse defined by the fuzzy set $f_i(t)$. If $F(t)$ consists of $f_i(t)$ (i = 1, 2,…), $F(t)$ is defined as a fuzzy time series on $Y(t)$ (t =…, 0, 1, 2,…) [14][15].

Definition 2: Let $F(t-1) = A_i$ and $F(t) = A_j$. The relationship between two consecutive observations, $F(t)$ and $F(t-1)$, referred to as a fuzzy logical relationship (FLR) (Song & Chissom, 1993a, 1993b), can be denoted by $A_i \rightarrow A_j$, where A_i is called the LHS (left-hand side) and A_j the RHS (right-hand side) of the FLR.

Definition 3: Let $F(t)$ be a fuzzy time series. If $F(t)$ is caused by $F_1(t-1)$, $F_2(t-1)$, $F_3(t-1)$,…, $F_g(t-1)$ (g is number of variables), then this fuzzy logical relationship(FLR) is defined by $F_1(t-1)$, $F_2(t-1)$, $F_3(t-1)$,…, $F_g(t-1) \rightarrow F_1(t)$ and it is called the one order multivariate fuzzy time series forecasting model, where $F_1(t)$, $F_2(t)$, $F_3(t)$,…, $F_g(t)$ are fuzzy time series, $F_1(t-1)$, $F_2(t-1)$, $F_3(t-1)$,…, $F_g(t-1)$ and $F_1(t)$ are called the current state and the next state, respectively [12].

This study applies artificial neural networks to establish the fuzzy relations, and also targets the multivariate problems. Hence, the multivariate fuzzy time series forecasting model is defined as follows:

Definition 4: Suppose $F_1(t)$, $F_2(t)$, $F_3(t)$ be three fuzzy time series, and $F_1(t-1) = A_p$, $F_2(t-1) = B_q$, $F_3(t-1) = C_r$, and $F_1(t) = A_s$. The multivariate FLR is defined as A_p, B_q, $C_r \rightarrow A_s$, where A_p, B_q, C_r are referred to as the LHS (left-hand side) and A_s as the RHS (right-hand side) of the multivariate FLR.

2.2 Feed forward neural networks

Since artificial neural networks (ANN) have introduced in 1950s, many various types of ANN models have been proposed. One of them is called as feed forward neural networks that have been used successfully in many studies [7]. It's a network structure consisting of number of Interconnect units by artificial neurons. Learning algorithms of ANN is to find all of the weights from the input generated by the corresponding desired output value for the specific task. Many different algorithms have been used for the determination of the optimal weights values. The most popularly used training method is the back propagation algorithm (BP) presented by Smith [13], it's also used in this study. This algorithm consists of adjusting all weights considering the error measure between the desired output and actual output [5].

A simple BPN structure is listed in Figure 1, it's composed by input layer, hidden layer, and output layer, respectively. In addition, the activation function of BPN, it determines the relationship between inputs and outputs of a network. The well-known activation functions are logistic, hyperbolic tangent, sine (or cosine) and the linear functions. Among them, logistic activation function is the most popular one [23], and another element of the BPN is the transfer function, it the output of BPN is depending on the transfer function used. Therefore, the BPN have the ability of learning and memory, through these functions and structure.

In this study, we use the BPN to training the fuzzy rules in the multivariate fuzzy time series, the difference with the data pretreatment is that all the variables through the fuzzy computing in the input variables of BPN, and not the normalization process. We will be explained our architecture of fuzzy-neural networks model in the third sections.

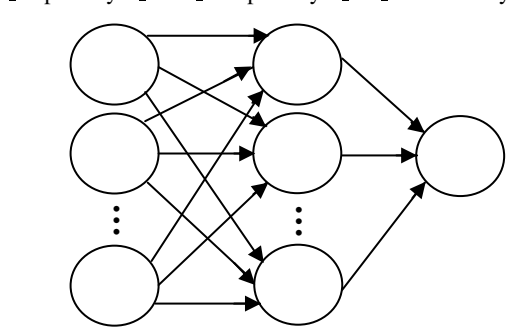


Figure 1. A feed forward back-propagation neural networks

2.3 Cumulative probability distribution

The CPDA have been successfully applied to the fuzzy time series forecast [2]. They propose a modified cumulative probability distribution approach to fuzzification the historical data. This method can be partitioned by mean and standard deviation (μ and σ) of the data, and the cumulative probability of normal distribution is used to determine the intervals. The procedure of CPDA is as follows:

Step1: To define the universe of discourse U.

Let $U = [D_{\min} - D_1,\ D_{\max} + D_2]$, where $D_{\min}$ and $D_{\max}$ are the minimum value and the maximum value of the training data, and D_1 and D_2 are two proper positive numbers. However, how to get D_1 and D_2 were determined was not explained.

Step2: To determine the length of intervals and build membership function.

The universe of discourse is partitioned into several intervals based on cumulative probability distribution, one critical decision is to determine the appropriate of the linguistic intervals, many previous studies have proposed different number of linguistic intervals [1][3][10][17]. In this study, we modified some equation of CPDA to build triangular fuzzy number. We defines n linguistic intervals from ratio of number of training data, the cumulative probability of each cut points are computed by

$$P = i/(n-1), \quad 1 \le i \le (n-2) \tag{1}$$

where i denotes the number of the cut points, and n denotes the total number of linguistic values (support intervals), respectively. The number of linguistic values can then be computed by the ratio of training data as

$$n = m \times \alpha, \quad \alpha = \{0.08, 0.09, 0.1\} \tag{2}$$

where m denotes the number of training data and α denotes the decision criteria of the linguistic intervals. The inverse of the normal cumulative distribution function (CDF) is computed with parameters μand σat the corresponding probabilities in P, where μand σ denote the mean and standard deviation of the data, respectively. The normal inverse function in terms of the normal CDF is defined as

$$x = F^{-1}(p|\mu,\sigma) = \{x : F(x|\mu,\sigma) = p\} \tag{3}$$

where $$p = F(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\prod}} \int_{-\infty}^{x} \frac{-(t-\mu)^2}{2\sigma^2}\, dt$$

The lower bound, midpoint, and upper bound as the triangular fuzzy number of each linguistic value can be computed according to the inverse of normal CDF, are listed in Table 1.

Step3: To fuzzification the historical data.

Establish fuzzy sets for the data and fuzzification the observations, shown as

$$\mu_{\tilde{L}_j}(x) \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & x > c \end{cases}, \quad j = 1, 2, \ldots, n \tag{4}$$

Table 1. The lower bound, midpoint, and upper bound as the triangular fuzzy number of each linguistic value.

Linguistic value	Lower bound(a)	Midpoint(b)	Upper bound(c)
A_1	D_{min}	D_{min}	$F^{-1}\left(\frac{1}{n-1}\middle\|\mu,\sigma\right)$
A_2	D_{min}	$F^{-1}\left(\frac{1}{n-1}\middle\|\mu,\sigma\right)$	$F^{-1}\left(\frac{2}{n-1}\middle\|\mu,\sigma\right)$
A_3	$F^{-1}\left(\frac{1}{n-1}\middle\|\mu,\sigma\right)$	$F^{-1}\left(\frac{2}{n-1}\middle\|\mu,\sigma\right)$	$F^{-1}\left(\frac{3}{n-1}\middle\|\mu,\sigma\right)$
⋮	⋮	⋮	⋮
A_{n-1}	$F^{-1}\left(\frac{n-3}{n-1}\middle\|\mu,\sigma\right)$	$F^{-1}\left(\frac{n-2}{n-1}\middle\|\mu,\sigma\right)$	D_{max}
A_n	$F^{-1}\left(\frac{n-2}{n-1}\middle\|\mu,\sigma\right)$	D_{max}	D_{max}

In this paper, we propose a multivariate fuzzy time series forecasting TAIEX, which factors were the VOL, TAIEX and TAIFEX, then define the universe of discourse and membership values at the three factors according to step 1 to step 3, respectively.

3. PROPOSED A MULTIVARIATE FUZZY TIME SERIES METHOD BASED ON NEURAL NETWORK

In this section, we propose a new method for the TAIEX forecasting is depicted in Figure 2. The algorithm step is shown as follows:

Step 1: Select feasible input variables according to the forecasting problems. Suppose the number of input variables is three fuzzy time series, $F_1(t)$, $F_2(t)$, $F_3(t)$.

Step 2: Define the universe of discourse and initially partition into n linguistic values, u_i. And apply the CPDA to generate intervals and fuzzify from the training data, respectively.

Step 3: Construct the fuzzy logic relationship(FLR) such as A_p, B_q, $C_r \rightarrow A_s$, for the training data.

Step 4: For neural network training, all the factors, where were the $F_1(t-1)$, $F_2(t-1)$ and $F_3(t-1)$ are taken as the inputs and the $F_1(t)$ is taken as the output, is depicted in Figure 3. The relation of input layer and output layer in the fuzzy-neural networks model, please refer the definition 3.

Step 5: Defuzzification each FLR to forecast, the forecasting value at time t can be calculated, shown as

$$F(t) = 0.5 \times (V(t-1) + D(A_j)) \tag{5}$$

where $V(t\text{-}1)$ denotes the actual value at time t-1 and $D(A_j)$ denotes defuzzified value and determined fuzzy rule, shown as

$$D(A_j) = \frac{a_{A_j} + b_{A_j} + c_{A_j}}{3} \tag{6}$$

where a_{Aj} , b_{Aj} *and* c_{Aj} denotes the lower bound, midpoint, and upper bound of the interval of A_j , respectively.

Step 6: To make forecasts for testing data according to the model trained by steps 1 to 5.

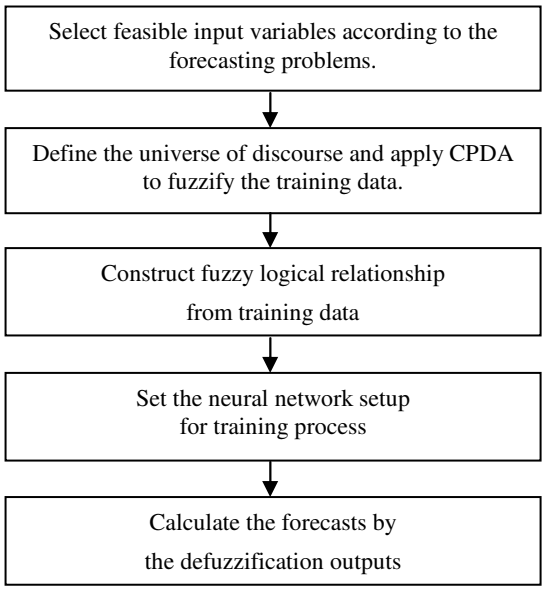


Figure 2. Forecasting processes of multivariate fuzzy time series method

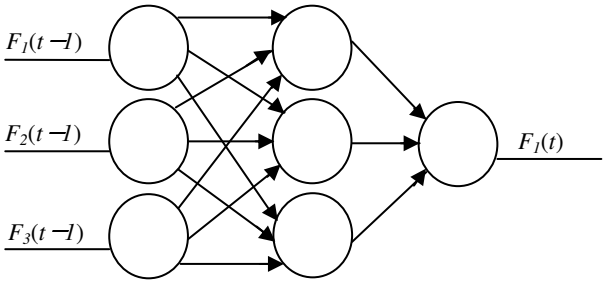


Figure. 3 The fuzzy-neural networks model

4. EXPERIMENTS

Since closing prices have been used in many previous studies [1] [8][10][11][17][18], they are also used in this study. The data are extracted from Taiwan Stock Exchange Corp (January 2001 - December 2004). In this section, we apply the proposed method to forecast the TAIEX. The proposed algorithm is now presented as follows:

Step 1: The stock of trading Volume, TAIEX and TAIFEX are taken as the inputs of neural network and the TAIEX is taken as the output of neural network, respectively. In addition, the data are divided into training and testing sets. Many studies have applied a convenient ratio to separate training (in-sample) from testing (out-of-sample) data [24]. This study follows the choice presented in [9][10][17]: the data covering January to October are used for training, while the data for November and December are used for testing. For example, the TAIEX of 2001 data covering the period 2001/1/2–2001/10/31, total 201 observations, are used for training, and the data covering the period 2001/11/01–2001/12/31, total 43 observations, are used for testing.

Step 2: Take the year 2001 as example. The minimal and maximal of TAIEX data in 2001 were 3446.26 and 6104.24, respectively. Hence, we set the universe of discourse as [3400, 6200] for the fuzzy set and apply the CPDA to generate intervals, the different ratio of number of training data as shown in Table 2.

Table 2. Different ratio of number of training data

Number of data	Ratio	Support intervals
201	0.08	16
201	0.09	18
201	0.1	20

We take the ratio is 0.1 for this example. Hence, the support intervals are 20, and linguistic values can be calculated, shown in Table 3. There has a problem is by using CPDA, that if the minimum and maximum of the probability distribution of value has overstep the stint of the universe of discourse U, we can sort all the linguistic intervals by ascending. Each fuzzy set, A_i, is defined by the intervals u_1, u_2, u_3, ..., u_{20}. Following [3][17], fuzzification data of the membership value whose maximum occurs at interval u_j, then the value of the data is fuzzified into A_j. For example, the TAIEX stock on 1/8/2001 was 4354.52, which is fuzzified to A_5 and the TAIEX stock on 2/8/2001 was 4490.19, which is fuzzified to A_6. In this step, we conduct a similar process for other factors, TAIFEX and VOL.

Table 3. Linguistic values of TAIEX stock by CPDA

Linguistic value	Lower bound(a)	Midpoint (b)	Upper bound(c)	$D(A_j)$
A_1	3,400.00	3,400.00	3,778.33	3,526.11
A_2	3,400.00	3,778.33	4,040.84	3,739.72
A_3	3,778.33	4,040.84	4,218.57	4,012.58
A_4	4,040.84	4,218.57	4,360.31	4,206.57
A_5	4,218.57	4,360.31	4,482.35	4,353.74
⋮	⋮	⋮	⋮	⋮
A_{19}	5,828.52	6,091.03	6,200.00	6,039.85
A_{20}	6,091.03	6,200.00	6,200.00	6,163.68

Step 3: We can construct the FLR, such as A_p, B_q, $C_r \rightarrow A_s$, as shown in Table 4, that the TAIEX, the TAIFEX, and the VOL of fuzzifed on 2001/1/2 are A_p, B_q, and C_r, respectively. And the TAIEX of fuzzifed on 2001/1/3 is A_s.. Hence, the FLR from 2001/1/2 to 2001/1/3 as shown as follows A_p, B_q, $C_r \rightarrow A_s$. Based on the FLR, the LHS (left-hand side) of the FLR are taken as inputs, respectively, and the RHS (right-hand side) of the FLR are taken as output, and the RHS is the forecast results of the neural network.

Step 4: For neural network training, one critical effect is to determine the appropriate of the neural network architecture [15]. Since in the literature, there are no general rules for determining the best architecture, much architecture must be examined for the correct results [6]. Hence, we set a architecture of neural network, that number of hidden layer is 1, number of nodes of hidden layer as 3, number of nodes of input layer as 3, number of nodes of output layer as 1, the epochs as 1000 and the learning rate as 0.5 to execute try and error, the architecture setup is show in Table 5.

Table 4. Linguistic values and fuzzy logic relationship

Data	TAIEX	Fuzzified TAIEX (A_p)	TAIFEX	Fuzzified TAIFEX (B_q)	VOL	Fuzzified VOL (C_r)	Fuzzy logic relationship
2001/1/2	4935.28	A_{11}	4892	B_{10}	64,870,815	C_{10}	$A_{11}, B_{10}, C_{10} \rightarrow A_{10}$.
2001/1/3	4894.79	A_{10}	4870	B_{10}	71,815,935	C_{11}	$A_{10}, B_{10}, C_{11} \rightarrow A_{13}$.
2001/1/4	5136.13	A_{13}	5121	B_{12}	89,357,810	C_{15}	$A_{13}, B_{12}, C_{15} \rightarrow A_{14}$.
2001/1/5	5295.53	A_{14}	5340	B_{15}	115,643,323	C_{18}	$A_{14}, B_{15}, C_{18} \rightarrow A_{13}$.
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 5. The neural network setup

Number of nodes of input layer	3
Number of hidden layer	1
Number of nodes of hidden layer	3
Number of nodes of output layer	1
Learning rate	0.5
Epochs	1000

Step 5: The output of training results is show in Table 6. For example, we can use equation (5) to forecast the TAIEX of 2001/01/04 from defuzzifed the FLR, that shown as

$$\begin{aligned} F(2001/01/04) &= 0.5 \times (V(2001/01/03) + D(A_{10})) \\ &= 0.5 \times (4894.79 + 4887.28) \\ &= 4891.03 \end{aligned}$$

where $D(A_j)$ can be calculated by equation (6). The forecasts for testing data are shown in Table 7.

Step 6: After training, we can add testing data into the existing training data to forecast and according to step 1 to step 5, respectively. In addition, the root mean square error (RMSE) is used to compare the performance of the proposed models and different methods, shown as equation (7) and the comparison models where are listed in Table 8.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n-1}(Actual(t) - Forecast(t))^2}{n-1}} \quad (7)$$

The results are showed in Table 8. It's proves that the proposed method has best predict ability than other method in years 2002, 2004, and overall performance.

When ratio is 0.08 or 0.09 can also have better predict ability than other method. Besides, if there are more support intervals, the performance are better than fewer support intervals. However, the result of year 2003 is different than other experiments. Maybe some are needed to more detailed experiments to explain this phenomenon.

5. Conclusions AND FUTURE WORK

In this study, we have proposed a multivariate fuzzy time series method based on fuzzy neural network. First, we apply the CPDA algorithm to generate different lengths of intervals in the universe of discourse. It's used the concept of probability distribution, that is used the pattern of the data themselves to be put into intervals rather than roughly putting them into intervals of static length. Second, we combine the neural network and multivariate fuzzy factors to forecast the TAIEX based on fuzzy time series.

Hence, this method, that is consider information-intensive to fuzzification historical data, and more objectivity to partitioned the universe of discourse, and it's usually to impact forecasting results of critical. In addition, using neural networks can be training the construction of fuzzy rules and predict performance through the learning mechanism. Based on the experimental results in section four, where reveal that the proposed method has better forecasting accuracies than some previous methods [10][17] for forecasting the TAIEX.

6. ACKNOWLEDGMENTS

The authors would like to thank the partially support sponsored by the National Science Council of Taiwan (R.O.C.) under the Grant NSC100-2410-H-324-003-.

Table 6. Training results of neural network for TAIEX from 2001/01/03 to 2001/10/31

Data	Actual TAIEX(t)	Input(A_{t-1})	Input(B_{t-1})	Input(C_{t-1})	Output(A_t)	Forecast TAIEX(t)
2001/1/3	4894.79	A_{10}	B_{10}	C_{11}	A_{11}	4958.68
2001/1/4	5136.13	A_{13}	B_{12}	C_{15}	A_{10}	4891.03
2001/1/5	5295.53	A_{14}	B_{15}	C_{18}	A_{14}	5207.82
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2001/10/30	3915.61	A_3	B_3	C_7	A_3	4038.84
2001/10/31	3903.49	A_2	B_2	C_7	A_3	3964.09

Table 7. Testing results of neural network for TAIEX from 2001/11/02 to 2001/12/31

Data	Actual TAIEX(t)	Input(A_{t-1})	Input(B_{t-1})	Input(C_{t-1})	Output(A_t)	Forecast TAIEX(t)
2001/11/2	3998.48	A_3	B_3	C_{10}	A_3	3971.13
2001/11/5	4080.51	A_3	B_3	C_{11}	A_3	4005.53
2001/11/6	4082.92	A_3	B_3	C_{14}	A_3	4046.54
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2001/12/28	5398.28	A_{15}	B_{15}	C_{19}	A_{17}	5497.88
2001/12/31	5551.24	A_{16}	B_{17}	C_{19}	A_{16}	5456.95

Table 8. Comparison of performance (RMSE) for different methods

Methods	Factors	Years:	2001	2002	2003	2004	Average
Huarng et al. [10]	TAIEX & NASDAQ		136.49	95.15	65.51	73.57	92.7
	TAIEX & Dow Jones		138.25	93.73	72.95	73.49	94.6
	TAIEX & M1b		133.26	97.1	75.23	80.01	96.4
	TAIEX & NASDAQ & Dow Jones		131.98	93.48	65.51	73.49	91.1
	TAIEX & NASDAQ & M1b		128.44	97.15	70.76	73.48	92.5
	TAIEX & NASDAQ & Dow Jones & M1b		124.02	95.73	70.76	72.35	90.7
Yu & Huarng [17]							
B_R model	TAIEX & TAIFEX		120*	77	54	85	84.0

B_NN model	TAIEX & TAIFEX	131	69	52*	61	78.3
B_NN_FTS model	TAIEX & TAIFEX	133	85	58	67	85.8
B_NN_FTS_S model	TAIEX & TAIFEX	130	80	58	67	83.8
The proposed method						
Ratio = 0.08	TAIEX & TAIFEX & VOL	124.9	74.5	59.4	61.6	80.1
Ratio = 0.09	TAIEX & TAIFEX & VOL	124.5	60.7	64.9	59.6	77.4
Ratio = 0.1	TAIEX & TAIFEX & VOL	123.5	56.4*	70	57*	76.7*

* best predict ability

7. REFERENCES

[1] J. R. Chang, L. Y. Wei, and C. H. Cheng. 2011. A Hybrid ANFIS model based on AR and volatility for TAIEX forecasting. *Applied Soft Computing*, 11(1):1388-1395.

[2] J. R. Chang, Y. T. Lee, S. Y. Liao, and C. H. Cheng. 2007. Cardinality-based Fuzzy Time Series for Forecasting Enrollments. (IEA/AIE 2007, Japan)H.G. Okuno and M. Ali (Eds.), *Lecture Notes in Artificial Intelligence*, 4570:735-744.

[3] S. M. Chen. 1996. Forecasting enrollments based on fuzzy time series. *Fuzzy sets and systems*, 81(3):311-319.

[4] C. H. Cheng, J. R. Chang, and C. A. Yeh. 2006. Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost. *Technological Forecasting and Social Change*, 73(5):524-542.

[5] Cichocki, A. and Unbehauen, R. 1993. Neural networks for optimization and signal processing. New York: John Wiley and Sons.

[6] Egrioglu, E., Aladag, C.H., and Gunay, S. 2008. A new model selection strategy in artificial neural networks. *Applied Mathematics and Computation*, 195(2):591-597.

[7] Gunay, S., Egrioglu, E., and Aladag, C.H. 2007. Introduction to Single Variable Time Series Analysis. Hacettepe University Press.

[8] K. Huarng, H. K.Yu. 2005. A Type 2 fuzzy time series model for stock index forecasting. *Physica A*, 353:445-462.

[9] K. Huarng, H. K. Yu. 2006. The application of neural networks to forecast fuzzy time series. *Physica A*, 363:481-491.

[10] K.H. Huarng, T. H.K. Yu, and Y.W. Hsu. 2007. A multivariate heuristic model for fuzzy time-series forecasting. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 37(4):836-846.

[11] Y. Y. Hsu, S. M. Tse, and B.Wu. 2003. A new approach of bivariate fuzzy time series analysis to the forecasting of a stock index. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(6):671-690.

[12] L. W. Lee, L. H. Wang, S. M. Chen, and Y. H. Leu. 2006. Handling forecasting problems based on two-factors high-order fuzzy time series. *IEEE Transactions on Fuzzy Systems*, 14(3):468-477.

[13] Smith, K.A. 2002. Neural networks in business: Techniques and applications. *Idea Group Publishing: Imprint Info Hershey*, PA.

[14] Q. Song, and B.S. Chissom. 1993a. Fuzzy time series and its models. *Fuzzy Sets and Systems*, 54(3):269-277.

[15] Q. Song, and B.S. Chissom. 1993b. Forecasting enrollments with fuzzy time series - Part I. *Fuzzy sets and systems*, 54(1):1-10.

[16] Q. Song, and B.S. Chissom. 1994. Forecasting enrollments with fuzzy time series - Part II. *Fuzzy Sets and Systems*, 62(1):1-8.

[17] H. K. Yu, and K.H. Huarng. 2008. A bivariate fuzzy time series model to forecast the TAIEX. *Expert Systems with Applications*, 34(4):2945-2952.

[18] T.H.K. Yu, and K.H. Huarng. 2010. Corrigendum to "A bivariate fuzzy series model to forecast the TAIEX". *Expert Systems with Applications*, 37(7):5529.

[19] Zadeh, L.A. 1965. Fuzzy Sets. *Information and Control*, 8(3):338-353.

[20] Zadeh, L.A. 1975a. The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8(3): 199-249.

[21] Zadeh, L.A. 1975b. The concept of a linguistic variable and its application to approximate reasoning-II. *Information Sciences*, 8(4):301-357.

[22] Zadeh, L.A. 1976. The concept of a linguistic variable and its application to approximate reasoning-III. *Information Sciences*, 9(1):43-80.

[23] G. Zhang, B.E. Patuwo, and Y.M. Hu. 1998. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14(1):35-62.

[24] G.P. Zhang. 2004. Business forecasting with artificial neural networks: an overview. *Idea Group Publishing: neural networks in business forecasting*, 1-22.

[25] Zurada, J.M. 1992. Introduction of artificial neural systems. St. Paul: West Publishing

ReClose Fuzz: Improved Automatic Summary Generation using Fuzzy Sets

Brent Wenerstrom
Computer Engineering and Computer Science
University of Louisville
Louisville, KY 40292 USA
bkwene01@louisville.edu

Rammohan Ragade
Computer Engineering and Computer Science
University of Louisville
Louisville, KY 40292 USA
ragade@louisville.edu

Mehmed Kantardzic
Computer Engineering and Computer Science
University of Louisville
Louisville, KY 40292 USA
mmkant01@louisville.edu

ABSTRACT
ReClose is an effective system for automatically generating web document summaries. It considers both query-biased and query-independent summary generation to create summaries using the best of both worlds. This approach has already been shown to provide summaries which lead to 20% more accurate user expectations compared to Google generated summaries [12]. We now show improvements to this system using both fuzzy sets and fuzzy numbers.

Keywords
fuzzy data mining, automatic summarization

1. INTRODUCTION
Search engines have become ubiquitous in internet usage. According to Pew as much as 92% of adult Internet users use a search engine to find information [7]. Each user that makes an internet search will be shown similar search result pages depending on the search engine chosen. A majority of search engines provide a linear list of web pages.

Each web pages is summarized by three main characteristics: the web page's title, a short text summary and the URL of the web page. Using only this information, each user must determine the worth of this web page to their information task. Users have two choices for each entry in the list of web pages in the search results. Either they can click through to the web page for more information or they can skip the result and move on. When a user successfully selects relevant web pages to click through, they are rewarded with a web page that gets them one step closer to the information they seek. Very often, users make wrong decisions due to poor summaries or wrong interpretations of the summary information provided.

The first possible mistake is that users can skip over relevant results. This is only a small hinderance if there are other relevant web documents in the search results. Occasionally, users will skip over the one relevant document in the search results and will therefore not be able to find any relevant documents for their search.

Of more concern generally is that users click on irrelevant web pages. This results in users wasting their time viewing a web page that contains no new information for the information task. This can lead to disappointment with the search results, running out of time or patience to find the relevant search results that remain on the result page, or even provide distractions that lower the productivity of the user.

A more accurate summary of a web page will lead more users to make correct decisions about clicking on a search result. This will mean more positive feelings towards a search engine and more successful searches leading people to the information they seek.

2. RELATED WORKS
Information retrieval has found improvements through the adoption of fuzzy systems. Bordogna and Pasi[2] extended the Boolean retrieval model using fuzzy sets which improved the expression power of the query language. Herrera-Viedma [4] added the mapping of certain terms in information retrieval to fuzzy ordinals to improve the user interaction with the system. Srinivasan et al. [10] explored the use of rough and fuzzy sets to expand query terms in an information retrieval environment.

There have also been several uses of fuzzy approaches to text mining. Rodriues and Sacks [9] created an efficient hierarchical fuzzy clustering algorithm for text mining which scaled linearly with the number of documents. Arotaritei and Mitra[1] surveyed fuzzy web mining solutions. Found were fuzzy approaches to clustering, association rule mining, personalization and text mining.

3. RECLOSE SYSTEM
ReClose[13] is web document summary generation system. It combines two approaches to summary generation. The first is query-biased summarization that incorporates query text when selecting sentences for summarization. This approach often reveals the relevance of a document to a query through one or more sentences that include keyword(s) from the query. The second approach, query-independent summarization focuses on summarizing the essence of a document to give the big picture view of a document. These two summary strategies combined to make ReClose a successful summary engine.

3.1 Query-Biased Summarization
The query-biased portion of ReClose built on top of successful search engine strategies for summarization. It used the summaries from Google, Yahoo and Bing to train a linear regression model to rank sentences in a document according to the likelihood that such a sentence would be used by these search giants in summarization [12].

First each document was fragmented into sentences along with each of the summaries from Google, Yahoo and Bing. Sentences were matched up between the summaries and the documents. For

each summary sentence that a sentence matched, it would increase the target output by 1/3. If a sentence was found both in Yahoo's summary and Google's summary, then that sentence would have a target output of 2/3.

Each sentence was converted to a number of representative, numeric attributes. We do not model the words of the sentence directly but rather attributes such as the percentage of query keywords present in a sentence and whether a verb was found to be present in the sentence or not. Then we trained a linear regression model on the target output using the calculated attributes. The output of this linear regression model was used to rank sentences in each new document for use as part of a query-biased summary.

3.2 Query-Independent Summarization

As of 2012 nearly every search engine generates summaries in a query-biased fashion. However, sometimes just reading that one or two sentences that include query keywords can be misleading. That is why we have added a query-independent element to the summary. Query-independent summaries provide a look at the document as a whole. Thus users can determine if the document as a whole is relevant or not. We use a graph representation of a document to rank sentences within a document for a query-independent summary of the document [13].

Again we split documents into sentences. However, now each sentence becomes a node in a graph. Then we compute similarity between each pair of sentences. This is where we have an opportunity to use fuzzy sets to improve the implementation. Then we use this weighted graph to rank sentences. Sentences are ranked by how similar they are to the rest of the graph, or by closeness centrality [3]. The rational is that the sentence most similar to the rest of the graph, best represents the content of the graph. Closeness centrality is averaged with the order of the sentence in the document. Then we have a sentence ranking for the document that represents another way to provide a query-independent summary.

3.3 Combination

Once we have sentence rankings from our query-independent and query-biased modules, then we can combine these rankings into a single summary. A naive approach would average the two rankings. This would result in sentences that are neither good at showing the relevance of a document to a query or good at summarizing the document, but have a medium level ranking in both. Such a summary would improve nothing upon current summary generation techniques. Instead we create a two part summary, where each part (query-biased and query-independent) portions are clearly labeled for the user to see. We take the highest ranking sentences from each ranking, while avoiding redundancy in content to populate the two parts. The result is a summary that is the best of both worlds.

4. IMPROVEMENT: SENTENCE SIMILARITY USING FUZZY SETS

To compute closeness centrality a graph is constructed from each web document. A node represents a sentence in the document and edges represent the dissimilarity between each sentence. Currently the approach to computing dissimalarity is computed as 1 - similarity. The similarity score is computed using the Jaccard index [5] on a bag of word representations to each sentence. This is computed as follows:

$$Similarity(s_1, s_2) = \frac{|s_1 \cap s_2|}{|s_1 \cup s_2|}$$

where s_1 and s_2 are sets which each contain an element for each different word within the two sentences, sentences 1 and 2.

Using word overlap to compute similarity between sentences could be replaced with another, more complex approach which uses fuzzy sets proposed by Murad and Martin [6]. Murad focuses on word similarities that may not be present in a thesaurus as is often used as the approach to find similar words. Murad looks at words in similar context. A word that is used in the same context as another word will often have the same meaning. Murad's approach requires that first a document has all stop words removed based on a user defined stop word list. Then all words are stemmed using Porter stemming [8].

A number of steps are required to compute the fuzzy sets used to represent each word in a document. For each word present in a document two sets are created. The first set, $p(x)$, defines all words that preceded the given word. The second set, $s(x)$, defines all words that followed the given word. We then count frequencies. We count the number of times the word occurs in a document. Then we count the number of times that certain words precede and follow a word. This is used to determine the membership function for the fuzzy set to be used for calculating similarity.

The quick brown fox jumps over the lazy dog.
The quick brown cat jumps onto the active dog.
The slow brown fox jumps onto the quick brown cat.
The quick brown cat leaps over the quick brown fox.

Figure 1.Example sentences used to explain similarity[6].

Figure 1 provides examples from Murad and Martin [6] that are used to explain the word fuzzy sets. There are four sentences in this document in all. Next we consider the input to the membership algorithm as described by Murad and Martin in Table 1. The input to the algorithm are frequency counts of the combination of preceding and successing words. For example considering the example sentences in Figure 1 and the word "brown," the counts for brown would be the following:

- quick – brown – cat × 3
- quick – brown – fox × 2
- slow – brown – fox × 1

Now we may convert the frequencies listed above into fuzzy sets using the formulas shown in Tables 1 and 2 we get the following:

- (quick, cat) $= m_{cw}[0] = 1$
- (quick, fox) $= m_{cw}[1] = m_{cw}[0] - (f_{cw}[0] - f_{cw}[1]) \times \frac{1}{7} = 1 - (3-2) \times \frac{1}{6} = 0.833$
- (slow, fox) $= m_{cw}[2] = 0.833 - (2-1) \times \frac{1}{6} = 0.5$

All that is needed are the frequencies of context for a given term within the current document.

Table 1.Input table for computing word fuzzy sets[6].

f_{cw}:	array of frequency counts.
T:	total frequency count for this word $= \sum_{(P,S)} f_{cw}(P,S)$ where P and S are precedence and successor respectively.

Table 2.Steps to establish membership from frequencies [6].

m_{cw}:	array of memberships
1.	Sort frequency counts into decreasing order, $f_{cw}[0] \ldots f_{cw}[n-1]$ such that $f_{cwi} \geq f_{cwj}$ iff $i \geq j$
2.	Set the membership corresponding to maximum count, $m_{cw}[0] = 1$
3.	for $i = 1 \ldots n-1$, i.e., for each remaining frequency count $m_{cw}[i] = m_{cw}[i-1] - (f_{cw}[i-1] - f_{cw}[i]) \times \frac{1}{T}$

Next we consider the computation of word probabilities. We can figure out the point semantic unification of two frequency distribution to determine how similarly two words are used. The higher the probability the more similar the contexts of the two words. To compute this semantic unification we need to calculate the membership of the first word using f_{cw1} and multiply it by the frequency of the corresponding element for the second term using f_{cw2}. This process is shown with the inputs to this process in Table 3 and the algorithm in Table 4.

Table 3.Input table for computing point semantic unifcation of two frequency distributions[6].

m_{cw1}:	array of memberships.
f_{cw2}:	array of frequency counts.
T_{cw2}:	total frequency count for this word: $\sum_{(P,S)} f_{cw2}(P,S)$ where P and S are precedence and successor respectively.

Consider some the fuzzy context set for some imagined word "grey" as the following:

- (quick, cat) = 1 with a count of 5 times.
- (slow, fox) = 0.75 with a count of 3 times.

Table 4.Steps of point semantic unification [6].

1.	Convert f_{cw1} to m_{cw1} using the steps outlined in Table 2.
2.	Calculate the sum of m_{cw1} multiply by f_{cw2} for the common elements giving the point semantic unification for two frequency distributions.
3.	To comput the asymmetric probability, simply reverse the calculation in steps 1 and 2.

We shall continue the example begun above for the word "brown". We would calculate the semantic unification of "brown" and "grey" as follows:

- $\Pr(brown|grey) = (m_{cw1}^{(quick,cat)} \times f_{cw2}^{(quick,cat)} + m_{cw1}^{(slow,fox)} \times f_{cw2}^{(slow,fox)}) \frac{1}{T_2} = \frac{1 \times 3 + 0.5 \times 3}{8} = 0.8125$
- $\Pr(grey|brown) = \frac{1 \times 3 + 0.75 \times 1}{8} = 0.625$

According to the above calculations the word "brown" could replace the word "grey" in the provided context distributions with a probability of 0.8125 and "grey" could replace "brown" with a chance of 0.625.

Lastly we combine the word semantic unification scores to create a sentence similarity metric [6] as shown below:

$$\sum_{i \in sent1} \sum_{j \in sent2} f_i \Pr(w_i|w_j) f_j$$

where f is the relative frequency of a word within a sentence. The $\Pr(w_i|w_j)$ is the semantic unification probability discussed earlier. When all of the semenatic word similarities are combined then we come up with a single similarity score for two sentences. These similarities can be used in place of the previous measure to compute possibly more attractive rankings of the summarizability of sentence for a particular document using closeness centrality. The same metrics could be used to measure the quality of the results as before.

5. EXPERIMENTAL RESULTS

Our experimental approach follows a similar pattern to that found in [11]. We downloaded some 229 news stories from CNN.com. These news stories covered a number of topics from world news, to politics to sports. They were downloaded during January and February in 2010. Each of the downloaded stories contained a section called "Story Highlights." This highlights section contained 3-4 bullet points highlighting key parts of the news story. These bullet points were human generated. Sometimes they contained direct quotes, and other times summarized multiple paragraphs in a single sentence. These highlights were used as the gold standard in measuring the effectiveness of ranking sentences for summary inclusion.

We measure our results using ROUGE-1. ROUGE-1 looks at the number of unigrams that co-occur both in the gold standard summary and the automatically produced summary. ROUGE-1 is calculated as follows:

$$ROUGE-1 = \frac{\sum_{gram_i \in S_1 \wedge S_2} Count(gram_i)}{\sum_{gram_i \in S} Count(gram_i)}$$

Using this CNN data set we set up an experiment in which we tested under which setup we could obtain the best ROUGE-1 scores. In this case we are testing closeness centrality using word overlap to measure sentence distances or fuzzy sets. We additionally test mixing sentence order with the rankings of closeness centrality.

Table 5. ROUGE-1 comparisons of distance metrics for closeness centrality.

Distance Metric for Closeness Centrality ranker	Alone	Combined with Sentence Order
Fuzzy Sets	0.477	0.502
Word Overlap	0.461	0.501

Since the fuzzy set approach was a similarity measure, we transformed it to distance by taking the inverse or 1/similarity. The results show that using fuzzy sets were able to achieve higher ROUGE-1 scores. The difference is not statistically significant, but is promising.

6. CONCLUSION

We looked at the use of fuzzy sets for sentence comparisons. We found that we could obtain improved results based on sentence ranking in CNN news stories when using fuzzy sets for sentence comparisons.

7. REFERENCES

[1] Arotaritei, D. and Mitra, S. 2004. Web mining: a survey in the fuzzy framework. *Fuzzy Sets and Systems*, 148(1):5 - 19.

[2] Bordogna, G. and Pasi, G. 1993. A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *Journal of the American Society for Information Science*, 44(2):70-82.

[3] Freeman, L.C. 1979. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215-239, 1978-1979.

[4] Herrera-Viedma, E. 2001. Modeling the retrieval process for an information retrieval system using an ordinal fuzzy linguistic approach. *Journal of the American Society for Information Science and Technology*, 52(6):460-475.

[5] Jaccard, P. 1901. Étude comparative de la distribution oraledansune portion des Alpes et des Jura. *Bulletin del la SociétéVaudoise des Sciences Naturelles*, 37:547-579.

[6] Murad, M. and Martin, T. 2007. Similarity-based estimation for document summarization using fuzzy sets. *International Journal of Computer Science and Security (IJCSS)*, 1(4):1-12, December 2007.

[7] Pew Research Center. Search and email still top the list of most popular online activities. `http://www.pewinternet.org/Reports/2011/Search-and-email/Report.aspx`, August 2011.

[8] Porter, M. 1980. An algorithm for suffix stripping. *Program*, 14(3):130-137.

[9] M. E. S. M. Rodrigues and L. Sacks. A scalable hierarchical fuzzy clustering algorithm for text mining. In *Proceedings of the 5th International Conference on Recent Advances in Soft Computing*, 2004.

[10] Srinivasan, P., Ruiz, M.E., Kraft, D.H., and J. Chen. 2001. Vocabulary mining for information retrieval: rough sets and fuzzy sets. *Information Processing & Management*, 37(1):15 - 38.

[11] Svore, K.M., Vanderwende, L., and Burges, C.J. 2007. Enhancing single-document summarization by combining ranknet and third-party sources. In *Proceedings of the Joint Conference on Empirical Methods on Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 448-457.

[12] Wenerstrom, B. and Kantardzic, M. 2010. Sentence ranking for search document summarizatin based on the wisdom of three search engines. In *Proceedings of the 11th International Conference on Internet Computing*, pages 277-283, July 2010.

[13] Wenerstrom, B. and Kantardzic, M. 2011. ReClose: Web page summarization combining summary techniques. *International Journal of Web Information Systems*, 7(4):333-359, 2011.

Observer Design for T-S Fuzzy Systems with Input Delay and Output Disturbance via an LMI Approach

Thai-Viet Dang
Department of Electrical Engineering, National Central University, JhongLi, 32001, Taiwan
dtviet-mtt@mail.hut.edu.vn

Wen-June Wang
Department of Electrical Engineering, National Central University, JhongLi, 32001, Taiwan
wjwang@ee.ncu.edu.tw

ABSTRACT
This paper presents a fuzzy observer synthesis for a T-S fuzzy system with bounded time-varying input delay and unknown output disturbance. First, an augmented fuzzy model is built by integrating the system state and the output disturbance into new variables. Next, based on the fuzzy model and Lyapunov theory, two main theorems are derived for particular and general cases of the fuzzy system, respectively. According to the main theorems and the linear matrix inequalities (LMIs) tools, the observer is synthesized to estimate the system state and the output disturbance simultaneously. Finally, an illustration with a numerical example is given to show that the synthesized observer is effective for the estimation of system's states and disturbance.

Keywords
Fuzzy observer, fuzzy system models, linear matrix inequalities (LMIs), input delay, output disturbance.

1. INTRODUCTION
Recently, the well-known Takagi–Sugeno (T–S) fuzzy model [1] has been paid considerable attention because of establishing stability analysis and stabilization for the unmodeled systems or un-clear model systems. Moreover, many systematic fuzzy controller and observer design methods (via state feedback and output feedback) have been developed based on the T-S model [2-8].

In practical control systems, state variables may be unavailable but are needed in the control design [16]. Therefore, the state estimation is required to deal with this situation [17]. Various approaches for the observer design have been proposed. Output feedback or observer-based control is proposed in [18-23] which pointed out a stabilizing output feedback controller provided that the stability of the controller and the exponential convergence of the observer are guaranteed by the Lyapunov method.

In view of possible time delays in practical systems, stability analysis and control synthesis of T–S fuzzy systems with state delay have been also addressed. The time-delay is always one of the reasons for breaking the stability or decreasing the quality of the dynamical systems. T–S fuzzy model was first used to deal with the stability analysis and control synthesis of nonlinear time delay systems in [8]. Consequently, the stability and stabilization of dynamical systems with time-delay have become important research issues in the control engineering areas [9-12].

Furthermore, the disturbance always exists with various forms in practical systems. In complex system models, the perturbations and modeling errors can be merged into as the disturbance. In general, the output disturbance is amplified in the conventional observer when the output disturbance exists in the considered systems [13] and [14]. Furthermore, most of the existing fuzzy observer approaches are not particularly suitable for handling fuzzy systems with output disturbances. However, the paper [15] brought out the solution of the observer design for the continuous T-S fuzzy system with output disturbance.

Motivated by the above concerns, the problem to be handled in the paper is the state and disturbance estimation subjected to the existence of unknown output disturbance in the continuous time-delay T-S fuzzy system. This work proposes a novel fuzzy observer to accurately estimate the system state and the output disturbance simultaneously. First, an augmented model is generated for the T-S fuzzy system. Then, based on Lyapunov theory, two main theorems are derived for particular and general cases of fuzzy systems, respectively. In each main theorem, the conditions are proposed under which the fuzzy observer is synthesized to estimate the system state and the output disturbance simultaneously. Furthermore, the condition is transformed into linear matrix inequalities (LMIs) [24] and [25] to find the parameters in the observer. According to the main theorem, a methodical procedure of the fuzzy observer synthesis is summarized. Finally, the observer is illustrated effectively by a numerical example.

The remainder of this paper is organized as follows. Several notations and problem statement are presented in Section 2, whereas Section 3 provides the main result and describes the observer design steps. A numerical example is presented in Section 4 to demonstrate the effectiveness of the proposed observer. Finally, some conclusions are given in Section 5.

2. PROBLEM DESCRIPTION
A T–S fuzzy dynamic model with input delay described by IF–THEN rules is considered, which represents the local linear input-output relation from a nonlinear uncertain system. The i^{th} rule of the fuzzy model is shown

Plant Rule i

IF θ_1 is μ_{i1} and … and θ_s is μ_{is}, THEN

$$\begin{aligned} \dot{x}(t) &= A_i x(t) + B_i u\left(t - \tau(t)\right) \\ y(t) &= C_i x(t), \end{aligned} \tag{1}$$

where $x(t) \in R^n$ is the unavailable state, $u(t) \in R^m$ is the control, $y(t) \in R^p$ the output, $\theta_j(t)$, $(j = 1, 2, ..., s)$ are the premise variables, and μ_{ij} $(i = 1, 2, ..., r; j = 1, 2, ..., s)$ are the fuzzy sets that are characterized by membership functions, r is the number of IF–THEN rules and s is the number of the premise

variables. $\tau(t)$ is an unknown bounded time-varying delay function, and the constant $\sigma > 0$ is the known upper bound of $\tau(t)$. The matrices A_i, B_i and C_i are with appropriate dimensions. The overall fuzzy model achieved from plant rules in (1) is given by

$$\begin{aligned}\dot{x}(t) &= \sum_{i=1}^{r} \beta_i(\theta(t))\{A_i x(t) + B_i u(t-\tau(t))\} \\ y(t) &= \sum_{i=1}^{r} \beta_i(\theta(t)) C_i x(t),\end{aligned} \tag{2}$$

The aforementioned model does not include the output disturbance. In practical cases, it is possible that the output contains an unknown disturbance

$$\begin{aligned}\dot{x}(t) &= \sum_{i=1}^{r} \beta_i(\theta(t))\{A_i x(t) + B_i u(t-\tau(t))\} \\ y(t) &= \sum_{i=1}^{r} \beta_i(\theta(t)) C_i x(t) + \omega(t),\end{aligned} \tag{3}$$

where $\omega(t) \in R^p$ stands for the unknown disturbance.

3. FUZZY OBSERVER SYNTHESIS

First, let us consider the particular case which has equal output matrices in (3) *i.e.,* $C_i = C$ for all *i*. Subsequently, this technique will be applied to the general case in which the system contains different output matrices.

3.1 The particular case

If the output matrices of all subsystems are equal. Then, (3) is rewritten as follows

$$\begin{aligned}\dot{x}(t) &= \sum_{i=1}^{r} \beta_i(\theta(t))\{A_i x(t) + B_i u(t-\tau(t))\} \\ y(t) &= Cx(t) + \omega(t).\end{aligned} \tag{4}$$

Rewrite (4) as the following augmented fuzzy system

$$\begin{aligned}\bar{E}\dot{\bar{x}}(t) &= \sum_{i=1}^{r} \beta_i(\theta)\{\bar{A}_i \bar{x}(t) + \bar{B}_i u(t-\tau) + \bar{N}\omega(t)\} \\ y(t) &= \bar{C}\bar{x}(t) = C_0 \bar{x}(t) + \omega(t),\end{aligned} \tag{5}$$

where $\bar{x}(t) = \begin{bmatrix} x(t) \\ \omega(t) \end{bmatrix}$, $\bar{E} = \begin{bmatrix} I_n & O_p \\ O_p & O_p \end{bmatrix}$, $\bar{A}_i = \begin{bmatrix} A_i & O_p \\ O_p & -I_p \end{bmatrix}$, $\bar{B}_i = \begin{bmatrix} B_i \\ O_{p\times m} \end{bmatrix}$,

$$\bar{N} = \begin{bmatrix} O_{n\times p} & I_p \end{bmatrix}^T,\ \bar{C} = \begin{bmatrix} C & I_p \end{bmatrix}, \text{ and } C_0 = \begin{bmatrix} C & O_p \end{bmatrix}. \tag{6}$$

However, in the real control problem, not all of the state variables are available. Therefore, it is necessary todesign an output feedback control law. We consider observer-based output feedback stabilization for system (5). The following observer rules are employed

Observer rule i :

IF θ_1 is μ_{i1} and … and θ_s is μ_{is}, THEN

$$\begin{aligned}\bar{E}_n \dot{\xi}(t) &= \bar{A}_{ni}\xi(t) + \bar{B}_i u(t-\tau(t)) \\ \hat{\bar{x}}(t) &= \xi(t) + \bar{K}_n y(t),\end{aligned} \tag{7}$$

The defuzzified output of (7) is represented as follows

$$\begin{aligned}\bar{E}_n \dot{\xi}(t) &= \sum_{i=1}^{r} \beta_i(\theta(t))\{\bar{A}_{ni}\xi(t) + \bar{B}_i u(t-\tau(t))\} \\ \hat{\bar{x}}(t) &= \xi(t) + \bar{K}_n y(t),\end{aligned} \tag{8}$$

Choose the control law as follow: $u(t) = \sum_{i=1}^{r} \beta_i(\theta(t)) \bar{K}_i \hat{\bar{x}}(t)$, where $\bar{K}_i$ are the local control gains. Suppose $\beta_i(\theta(t)) = \beta_i(t)$ and $\beta_i(\theta(t-\tau)) = \beta_i(t-\tau)$.

Lemma 1: [26] For any constant matrix $D > 0$, any scalars *a* and *b* with $a < b$, and a vector function $x(t):[a,b] \to R^n$ such that the integrals concerned are well defined

$$\left[\int_a^b x(s)ds\right]^T D\left[\int_a^b x(s)ds\right] \le (b-a)\int_a^b x^T(s)Dx(s)ds. \tag{9}$$

Lemma 2: [27] Let $a \in R^n$, $b \in R^m$ and $G \in R^{m\times n}$. Then, for any matrices *X*, *Y*, and *Z* with appropriate dimensions, the following inequality hold

$$-2a^T Gb \le \begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} X & Y-G \\ * & Z \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \tag{10}$$

where *X*, *Y*, and *Z* satisfy $\begin{bmatrix} X & Y-G \\ * & Z \end{bmatrix} \ge 0$.

Substitute into the (8), we obtain (11) as follows

$$\begin{aligned}&\bar{E}_n \dot{\hat{\bar{x}}}(t) - \bar{E}_n \bar{K}_n \bar{C}\dot{\bar{x}}(t) \\ &= \sum_{i=1}^{r} \beta_i(t)\{\bar{A}_{ni}(\hat{\bar{x}}(t) - \bar{K}_n C_0 \bar{x}(t) - \bar{K}_n \omega(t)) + \bar{B}_i u(t-\tau)\}\end{aligned} \tag{11}$$

Let $\bar{e}(t) = \bar{x}(t) - \hat{\bar{x}}(t)$, (5) can be rewritten as follows

$$\begin{aligned}\bar{E}\dot{\bar{x}}(t) &= \sum_{i,j=1}^{r} \beta_i(t)\beta_j(t-\tau) \\ &\times\{\bar{A}_i(\bar{x}(t) + \bar{B}_i\bar{K}_j\bar{x}(t-\tau) - \bar{B}_i\bar{K}_j\bar{e}(t-\tau) + \bar{N}\omega(t))\}.\end{aligned} \tag{12}$$

Noted that $\bar{x}(t)-\bar{x}(\tau)=\int_{t-\tau}^{t}\dot{x}(\kappa)d\kappa$, (12) can be rewritten as

$$\bar{E}\dot{\bar{x}}(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\Big\{\big(\bar{A}_i+\bar{B}_i\bar{K}_j\big)\bar{x}(t) -\bar{B}_i\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{x}}(\kappa)d\kappa-\bar{B}_i\bar{K}_j\bar{e}(t-\tau)+\bar{N}\omega(t)\Big\}. \quad (13)$$

and (11) as the form below

$$\bar{E}_n\dot{\hat{\bar{x}}}(t)-\bar{E}_n\bar{K}_n\bar{C}\dot{\bar{x}}(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\Big\{\big(\bar{A}_{ni}+\bar{B}_i\bar{K}_j\big)\bar{x}(t) -\bar{A}_{ni}\big(\bar{K}_nC_0x(t)+\bar{K}_n\omega(t)\big)-\bar{B}_i\bar{K}_j\int_{t-\tau}^{t}\dot{\hat{\bar{x}}}(\kappa)d\kappa\Big\}. \quad (14)$$

Subtract (14) from (13) yields

$$\begin{aligned}&\big(\bar{E}+\bar{E}_n\bar{K}_n\bar{C}\big)\dot{\bar{x}}(t)-\bar{E}_n\dot{\hat{\bar{x}}}(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\\&\times\Big\{\big(\bar{A}_i+\bar{B}_i\bar{K}_j+\bar{A}_{ni}\bar{K}_nC_0\big)\bar{x}(t)-\big(\bar{A}_{ni}+\bar{B}_i\bar{K}_j\big)\hat{\bar{x}}(t)+\bar{N}\omega(t)\\&\quad+\bar{A}_{ni}\bar{K}_n\omega(t)-\bar{B}_i\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa-\bar{B}_i\bar{K}_j\bar{e}(t-\tau)\Big\}.\end{aligned} \quad (15)$$

Equilibrating parameters of equation (15) implies these equations as follows

$$\begin{aligned}&\bar{A}_i+\bar{B}_i\bar{K}_j+\bar{A}_{ni}\bar{K}_nC_0=\bar{A}_{ni}+\bar{B}_i\bar{K}_j\\&\bar{N}=-\bar{A}_{ni}\bar{K}_n\\&\big(\bar{E}+\bar{E}_n\bar{K}_n\bar{C}\big)=\bar{E}_n.\end{aligned} \quad (16)$$

From (16) we obtain one suitable set of $\big(\bar{A}_{ni},\ \bar{K}_n\ ,\bar{E}_n\big)$ as below

$$\bar{A}_{ni}=\begin{bmatrix}A_i & 0\\ -C & -I_p\end{bmatrix},\bar{K}_n=\begin{bmatrix}0\\ I_p\end{bmatrix}\text{ and }\bar{E}_n=\begin{bmatrix}I & 0\\ MC & M\end{bmatrix}, \quad (17)$$

where $M\in R^{n\times p}$ is a full-rank matrix. One obtains the error dynamic system

$$\bar{E}_n\dot{\bar{e}}(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\Big\{\big(\bar{A}_i+\bar{B}_i\bar{K}_j\big)\bar{x}(t) -\bar{B}_i\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{x}}(\kappa)d\kappa-\bar{B}_i\bar{K}_j\bar{e}(t-\tau)\Big\}, \quad (18)$$

with the aid of (17), the error dynamic system (18) becomes

$$\dot{\bar{e}}(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\Big\{\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)\bar{e}(t) -\bar{B}_{1i}\bar{K}_j\Big(\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\Big)-\bar{B}_{1i}\bar{K}_j\bar{e}(t-\tau)\Big\} \quad (19)$$

where $\bar{A}_{1i}=\begin{bmatrix}I & 0\\ MC & M\end{bmatrix}^{-1}\begin{bmatrix}A_i & 0\\ -C & Ip\end{bmatrix},\bar{B}_{1i}=\begin{bmatrix}I & 0\\ MC & M\end{bmatrix}^{-1}\bar{B}_i$ (20)

Theorem 1: Given scalar $\sigma>0$ and suppose $\bar{A}_{ni}$, $\bar{K}_n$ and $\bar{E}_n$ are defined in (17). The fuzzy observer (8) will estimate asymptotically the state and output disturbance of fuzzy system (5) if the following conditions hold.

There exist some matrices $\bar{P}>0$, $\bar{Q}>0$, $\bar{S}>0$, $\bar{X}_{1i}$, $\bar{X}_{2i}$, $\bar{Y}_{1i}$, $\bar{Y}_{2i}$, $\bar{Z}_{1i}$, $\bar{Z}_{2i}$ and $\bar{K}_i$ satisfying the following LMIs for i,j =1, 2,... r.

$$\begin{bmatrix}\bar{\Phi}_i & \bar{Y}_{2i}-\bar{P} & 0 & \bar{Y}_{1i}-\bar{P}\bar{B}_{1i}\bar{K}_j\\ * & \bar{Z}_{2i}-\bar{S} & 0 & 0\\ * & * & \sigma\bar{Q} & 0\\ * & * & * & \bar{Z}_{1i}-\frac{1}{\sigma}\bar{Q}\end{bmatrix}<0, \quad (21)$$

$$\begin{bmatrix}\bar{X}_{1i} & \bar{Y}_{1i}\\ * & \bar{Z}_{1i}\end{bmatrix}\geq 0, \quad (22)$$

and $$\begin{bmatrix}\bar{X}_{2i} & \bar{Y}_{2i}\\ * & \bar{Z}_{2i}\end{bmatrix}\geq 0, \quad (23)$$

where $\bar{\Phi}_i=\bar{P}\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)+\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)^T\bar{P}+\bar{X}_{1i}+\bar{X}_{2i}+\bar{S}$.

Here, symbol * denotes the transpose elements (matrices) for symmetric positions.

Proof: Let the Lyapunov candidate be $V=V_1+V_2+V_3$ with

$$V_1=\bar{e}^T(t)\bar{P}\bar{e}(t),\ V_2=\int_{t-\sigma}^{t}(s-t+\sigma)\dot{\bar{e}}^T(\alpha)\bar{Q}\dot{\bar{e}}(\alpha)d\alpha,$$

and $V_3=\int_{t-\sigma}^{t}\bar{e}^T(\alpha)\bar{S}\bar{e}(\alpha)d\alpha$.

Differentiating V along the trajectory of (19) yields

$$\begin{aligned}\dot{V}_1&=\dot{\bar{e}}^T(t)\bar{P}\bar{e}(t)+\bar{e}^T(t)\bar{P}\dot{\bar{e}}(t)\\&=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\Big\{\bar{P}\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)+\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)^T\bar{P}\Big\}\bar{e}(t)\\&\quad-2\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\bar{P}\bar{B}_{1i}\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\\&\quad-2\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\bar{P}\bar{e}(t-\tau).\end{aligned} \quad (24)$$

$$\dot{V}_2=\sigma\dot{\bar{e}}^T(t)\bar{Q}\dot{\bar{e}}(t)-\int_{t-\sigma}^{t}\dot{\bar{e}}^T(\alpha)\bar{Q}\dot{\bar{e}}(\alpha)d\alpha\leq\sigma\dot{\bar{e}}^T(t)\bar{Q}\dot{\bar{e}}(t) \quad (25)$$

$$-\int_{t-\tau}^{t}\dot{\bar{e}}^T(\alpha)\bar{Q}\dot{\bar{e}}(\alpha)d\alpha.$$

By using Lemma 1:

$$\int_{t-\tau}^{t}\dot{\bar{e}}^T(\alpha)\bar{Q}\dot{\bar{e}}(\alpha)d\alpha\leq\frac{1}{\sigma}\Big[\int_{t-\tau}^{t}\dot{\bar{e}}(\alpha)d\alpha\Big]^T\bar{Q}\int_{t-\tau}^{t}\dot{\bar{e}}(\alpha)d\alpha. \quad (26)$$

By using Lemma 2:

$$\begin{aligned}&-2\bar{e}^T(t)\bar{P}\bar{B}_{1i}\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\\&\leq\begin{bmatrix}\bar{e}(t)\\ \int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\end{bmatrix}^T\begin{bmatrix}\bar{X}_{1i} & \bar{Y}_{1i}-\bar{P}\bar{B}_{1i}\bar{K}_j\\ * & \bar{Z}_{1i}\end{bmatrix}\begin{bmatrix}\bar{e}(t)\\ \int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\end{bmatrix},\end{aligned} \quad (27)$$

and

$$-2\bar{e}^T(t)\bar{P}\bar{e}(t-\tau)\leq\begin{bmatrix}\bar{e}(t)\\ \bar{e}(t-\tau)\end{bmatrix}^T\begin{bmatrix}\bar{X}_{2i} & \bar{Y}_{2i}-P\\ * & \bar{Z}_{2i}\end{bmatrix}\begin{bmatrix}\bar{e}(t)\\ \bar{e}(t-\tau)\end{bmatrix}. \quad (28)$$

$$\dot{V}_3=\bar{e}^T(t)\bar{S}\bar{e}(t)-\bar{e}^T(t-\tau)\bar{S}\bar{e}(t-\tau). \quad (29)$$

Thus, from (24)~ (29), it yields

$$\dot{V} \le \sum_{i,j=1}^{r} \beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\Big\{\bar{P}\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)+\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)^T\bar{P}$$

$$+\bar{X}_{1i}+\bar{X}_{2i}+\bar{S}\Big\}\bar{e}(t)+\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\big[\bar{Y}_{1i}-\bar{P}\bar{B}_{1i}\bar{K}_j\big]$$

$$\times\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa+\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t)\big[\bar{Y}_{2i}-\bar{P}\big]\bar{e}(t-\tau)$$

$$+\sigma\dot{\bar{e}}^T(t)\bar{Q}\dot{\bar{e}}(t)+\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\left[\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\right]^T\left(\bar{Z}_{1i}-\frac{1}{\sigma}\bar{Q}\right)$$

$$\times\left[\int_{t-\tau}^{t}\dot{\bar{e}}(\kappa)d\kappa\right]+\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{e}^T(t-\tau)\big(\bar{Z}_{2i}-\bar{S}\big)\bar{e}(t-\tau).$$

$$\dot{V} \le \sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)\bar{z}^T(t)$$

$$\times\begin{bmatrix}\bar{\Phi}_i & \bar{Y}_{2i}-\bar{P} & 0 & \bar{Y}_{1i}-\bar{P}\bar{B}_{1i}\bar{K}_j\\ * & \bar{Z}_{2i}-\bar{S} & 0 & 0\\ * & * & \sigma\bar{Q} & 0\\ * & * & * & \bar{Z}_{1i}-\dfrac{1}{\sigma}\bar{Q}\end{bmatrix}\bar{z}(t)<0, \tag{30}$$

where $\bar{z}^T(t)=\left[\bar{e}^T(t)\quad \bar{e}^T(t-\tau)\quad \dot{\bar{e}}^T(t)\quad \int_{t-\tau}^{t}\dot{\bar{e}}^T(\kappa)d\kappa\right]$.

Finally, the proof is successfully completed.

3.2 The general case

The output matrices of all subsystems may not be equal, *i.e.*, $C_i \neq C_j$, $i \neq j$. Then, the T-S fuzzy system can be rewritten as

$$\dot{x}(t)=\sum_{i=1}^{r}\beta_i(\theta(t))\big\{A_ix(t)+B_iu(t-\tau(t))\big\}$$
$$y(t)=Cx(t)+\sum_{i=1}^{r}\beta_i(t)(C_i-C)x(t)+\omega(t). \tag{31}$$

where C is the matrix chosen from the set C_1, C_2..., C_r.

Let $\omega_0(t)=\sum_{i=1}^{r}\beta_i(t)(C_i-C)x(t)+\omega(t)$, (31) becomes

$$\dot{x}(t)=\sum_{i=1}^{r}\beta_i(t)\big\{A_ix(t)+B_iu(t-\tau(t))\big\}$$
$$y=Cx(t)+\sum_{i=1}^{r}\beta_i(t)(C_i-C)x(t)+\omega_0(t). \tag{32}$$

Similar to (5), (32) can be rewritten as

$$\bar{E}\dot{\bar{x}}_0(t)=\sum_{i=1}^{r}\beta_i(t)\big\{\bar{A}_i\bar{x}_0(t)+\bar{B}_iu(t-\tau)+\bar{N}\omega_0(t)\big\} \tag{33}$$
$$y(t)=\bar{C}\bar{x}_0(t)=C_0\bar{x}_0(t)+\omega_0(t),$$

where $\bar{x}_0^T=\big[x(t)\quad \omega_0(t)\big]$ and the state-space system coefficients $\bar{A}_i$, $\bar{B}_i$, $\bar{N}$, $\bar{C}$, C_0 and $\bar{E}$ are defined as the same as those in (6). Therefore, we can design the T-S fuzzy observer to estimate $x(t)$ and $\omega_0(t)$ of system (33) by using the approach obtained in Theorem 1. Furthermore, we can estimate $x(t)$ and $\omega(t)$ by the following dynamic system:

$$\bar{E}_n\dot{\hat{\xi}}(t)=\sum_{i=1}^{r}\beta_i(\theta(t))\big\{\bar{A}_{ni}\xi(t)+\bar{B}_iu(t-\tau(t))\big\} \tag{34}$$
$$\hat{\bar{x}}_0(t)=\xi(t)+\bar{K}_ny(t)$$
$$\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}(t)\end{bmatrix}=\begin{bmatrix}I_n & 0\\ \sum_{i=1}^{r}\beta_i(t)(C_i-C) & I_p\end{bmatrix}^{-1}\hat{\bar{x}}_0(t)$$

where $\hat{\bar{x}}_0(t)$ is the estimate of $\bar{x}_0(t)$. Therefore, it yields

$$\dot{\bar{e}}_0(t)=\sum_{i,j=1}^{r}\beta_i(t)\beta_j(t-\tau)$$
$$\times\left\{\big\{\big(\bar{A}_{1i}+\bar{B}_{1i}\bar{K}_j\big)\bar{e}_0(t)-\bar{B}_{1i}\bar{K}_j\int_{t-\tau}^{t}\dot{\bar{e}}_0(\kappa)d\kappa-\bar{B}_{1i}\bar{K}_j\bar{e}_0(t-\tau)\big\}\right\} \tag{35}$$

where $\bar{A}_{1i}$, $\bar{B}_{1i}$ and $\bar{K}_i$ are the same as those in the particular case.

Theorem 2: Consider the fuzzy system (3) and fuzzy observer (34). Suppose $\bar{A}_{ni}$, $\bar{K}_n$ and $\bar{E}_n$ are defined in (17) then, the estimation errors (35) of state and the output disturbance converge to zero asymptotically, if the conditions in Theorem 1 hold simultaneously.

Proof:

Because the forms of fuzzy system (33) and fuzzy observer (34) are similar to the forms of fuzzy system (5) and fuzzy observer (8), respectively, based on Theorem 1, if conditions in Theorem 1 hold simultaneously,

$$\lim\left\{\begin{bmatrix}x(t)\\ \omega_0(t)\end{bmatrix}-\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}_0(t)\end{bmatrix}\right\}=0, \tag{36}$$

Since $\beta_i(t)$ is bounded, it obtains

$$\lim_{t\to\infty}\begin{bmatrix}I_n & 0\\ \sum_{i=1}^{r}\beta_i(t)(C_i-C) & I_p\end{bmatrix}^{-1}\left\{\begin{bmatrix}x(t)\\ \omega_0(t)\end{bmatrix}-\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}_0(t)\end{bmatrix}\right\}=0. \tag{37}$$

and $\begin{bmatrix}x(t)\\ \omega(t)\end{bmatrix}-\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}(t)\end{bmatrix}=\begin{bmatrix}x(t)\\ \omega(t)\end{bmatrix}-\begin{bmatrix}I_n & 0\\ \sum_{i=1}^{r}\beta_i(\theta)(C_i-C) & I_p\end{bmatrix}^{-1}\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}_0(t)\end{bmatrix}$

$$=\begin{bmatrix}I_n & 0\\ \sum_{i=1}^{r}\beta_i(\theta)(C_i-C) & I_p\end{bmatrix}^{-1}$$

$$\times\left\{\begin{bmatrix}I_n & 0\\ \sum_{i=1}^{r}\beta_i(\theta)(C_i-C) & I_p\end{bmatrix}\begin{bmatrix}x(t)\\ \omega(t)\end{bmatrix}-\begin{bmatrix}\hat{x}(t)\\ \hat{\omega}_0(t)\end{bmatrix}\right\} \tag{38}$$

Based on (37) and (38), it yields (36). The proof is completed.

Remark 1: Consider the LMIs as the form of (39), (40), and (41) as below

$$\begin{bmatrix}\bar{\Phi}_i & Y_{2i}-P & 0 & Y_{1i}-B_{1i}\bar{K}_j & 0\\ * & Z_{2i}-S & 0 & 0 & 0\\ * & * & \sigma Q & 0 & 0\\ * & * & * & Z_{1i}-\dfrac{1}{\sigma}Q & 0\\ * & * & * & * & -I\end{bmatrix}<0, \tag{39}$$

$$\begin{bmatrix} X_{1i} & Y_{1i} \\ * & Z_{1i} \end{bmatrix} \geq 0, \quad (40)$$

and $$\begin{bmatrix} X_{2i} & Y_{2i} \\ * & Z_{2i} \end{bmatrix} \geq 0, \quad (41)$$

where $\Phi_i = \bar{A}_{1i}P + P\bar{A}_{1i}^T + \bar{B}_{1i}F_j + F_j\bar{B}_{1i}^T + X_{1i} + X_{2i} + S.$

Suppose $\bar{\Phi}_i = \bar{P}\left(\bar{A}_{1i} + \bar{B}_{1i}K_j\right) + \left(\bar{A}_{1i} + \bar{B}_{1i}K_j\right)^T \bar{P} + \bar{X}_{1i} + \bar{X}_{2i} + \bar{S},$

$\bar{X}_{ij} = \bar{P}X_{ij}\bar{P},\ \bar{Y}_{ij} = \bar{P}Y_{ij}\bar{P},\ \bar{Z}_{ij} = \bar{P}Z_{ij}\bar{P},\ \bar{S} = \bar{P}S\bar{P},\ \bar{Q} = \bar{P}Q\bar{P}.$ (42)

Moreover, the feedback gain matrices K_i are supposed as $K_i = F_iP^{-1}$ and $\bar{P} = P^{-1}$. We will prove (39), (40), and (41) to be equivalent to (21), (22), and (23).

Proof:

Pre- and post-multiplying (39) by $diag\left[\bar{P} \quad \bar{P} \quad \bar{P} \quad \bar{P} \quad I\right]$ and then, applying the Schur complement to yield a new form (21). Next, multiply both sides of (40) and (41) with $diag\left[\bar{P} \quad \bar{P}\right]$ to obtain the equivalent LMIs (22) and (23). So now, solve the LMIs (39), (40), and (41) to find X, S, Q and F_i. Finally, obtain feedback gains $\bar{K}_i$, $\bar{P}$, $\bar{Q}$ and $\bar{S}$.

Synthesis procedure for the observer:

Step 1: Set up the fuzzy model (3) for the nonlinear system. Then, make the augmented fuzzy system (33).

Step 2: Set up the fuzzy observer (34) for T-S fuzzy system (33) by choosing matrix *M* to guarantee (17) hold.

Step 3: Choose the scalar $\sigma > 0$.

Step 4: Solve the LMIs (39), (40), and (41) to find P, S, Q and F_i.

Step 5: If the LMIs (39), (40), and (41) are infeasible solutions with the aids of LMI-Toolbox of MATLAB, go back Step 3.

Step 6: If the LMIs (39), (40), and (41) are feasible solutions, obtain feedback gains $\bar{K}_i$, $\bar{P}$, $\bar{Q}$ and $\bar{S}$.

Finally, the fuzzy observer is completely synthesized.

4. A NUMERICAL EXAMPLE

Step 1: Consider a nonlinear system described by the following T-S fuzzy system

Rule 1. If $y_1^2(t)$ is μ_1, then $\dot{x}(t) = A_1x(t) + B_1u(t-\tau)$
$y(t) = C_1x(t).$

Rule 2. If $y_1^2(t)$ is μ_2, then $\dot{x}(t) = A_2x(t) + B_2u(t-\tau)$
$y(t) = C_2x(t).$

where $A_1 = \begin{bmatrix} -1 & -2 \\ 2 & -1 \end{bmatrix}$, $A_2 = \begin{bmatrix} -2 & 1 \\ -0.5 & -1 \end{bmatrix}$, $B_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, B_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$,

$C_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ and $C_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Step 2: Choose $M = \begin{bmatrix} -0.01 & -0.02 \\ 0 & -0.01 \end{bmatrix}$ and $C = C_1$ to $\begin{bmatrix} I & 0 \\ MC & M \end{bmatrix}^{-1}$ exist. Then, obtain one suitable set of $\left(\bar{A}_{ni}, \bar{K}_n, \bar{E}_n\right)$ in (17).

$$\bar{A}_{ni} = \begin{bmatrix} -1 & -2 & 0 & 0 \\ 2 & -1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 \end{bmatrix}, \quad K_n = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and $$E_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -0.03 & -0.02 & -0.01 & -0.02 \\ -0.01 & -0.01 & 0 & -0.01 \end{bmatrix}.$$

Step 3: Choose $\sigma = 0.01$.

Step 4: Solve the LMIs (50), (51), and (52) obtain

$$P = \begin{bmatrix} 0.0001\times10^7 & -0.0001\times10^7 & 0 & 0 \\ -0.0001\times10^7 & 0.0001\times10^7 & -0.0001\times10^7 & 0 \\ -0.0001\times10^7 & 0.0001\times10^7 & -4.8757\times10^7 & 1.6155\times10^7 \\ 0 & 0 & 1.6155\times10^7 & 1.6264\times10^7 \end{bmatrix},$$

$$S = \begin{bmatrix} 0.0001\times10^7 & -0.0001\times10^7 & 0 & 0 \\ -0.0001\times10^7 & 0.0001\times10^7 & -0.0001\times10^7 & 0 \\ 0 & -0.0001\times10^7 & 4.3249\times10^7 & -0.0032\times10^7 \\ 0 & 0 & -0.0032\times10^7 & 4.3195\times10^7 \end{bmatrix},$$

$F_1 = \begin{bmatrix} -0.2867\times10^{-13} & -0.0398\times10^{-13} & -0.0010\times10^{-13} & -0.0002\times10^{-13} \end{bmatrix}$,

$F_2 = \begin{bmatrix} -0.0761\times10^{-13} & -0.2945\times10^{-13} & -0.0001\times10^{-13} & 0 \end{bmatrix}$.

$$Q = \begin{bmatrix} 0.1441\times10^{-14} & -0.0054\times10^{-14} & -0.0003\times10^{-14} & 0 \\ -0.0054\times10^{-14} & 0.1452\times10^{-14} & 0 & -0.0001\times10^{-14} \\ -0.0003\times10^{-14} & 0 & 0.1124\times10^{-14} & 0 \\ 0 & -0.0001\times10^{-14} & 0 & 0.1125\times10^{-14} \end{bmatrix},$$

Step 5: Obtain feedback gains

$$\bar{P} = \begin{bmatrix} 4.6369\times10^{12} & 0.0775\times10^{12} & 0 & 0 \\ 0.0775\times10^{12} & 4.6817\times10^{12} & 0.0001\times10^{12} & 0 \\ 0 & 0.0001\times10^{12} & 0.0001\times10^{12} & -0.0001\times10^{12} \\ 0 & 0 & -0.0001\times10^{12} & 0.0001\times10^{12} \end{bmatrix},$$

$$\bar{S} = \begin{bmatrix} 4.8198\times10^{12} & 0.0527\times10^{12} & 0 & 0 \\ 0.0527\times10^{12} & 4.7442\times10^{12} & -0.0001\times10^{12} & 0 \\ 0 & -0.0001\times10^{12} & 0.0001\times10^{12} & -0.0001\times10^{12} \\ 0 & 0 & -0.0001\times10^{12} & 0.0001\times10^{12} \end{bmatrix},$$

$K_1 = \begin{bmatrix} -0.1333 & -0.0209 & 0.0001 & 0 \end{bmatrix}, K_2 = \begin{bmatrix} -0.0376 & -0.1384 & -0.0001 & 0 \end{bmatrix}$,

$$\bar{Q} = \begin{bmatrix} 3.1027\times10^{10} & 0.2207\times10^{10} & 0 & 0 \\ 0.2207\times10^{10} & 3.1867\times10^{10} & 0.0001\times10^{10} & 0 \\ 0 & 0.0001\times10^{10} & 0.0001\times10^{10} & 0.0001\times10^{10} \\ 0 & 0 & 0.0001\times10^{10} & 0.0001\times10^{12} \end{bmatrix}.$$

Suppose *u(t) = sin(t)* and $\omega(t) = \begin{bmatrix} \omega_1(t) & \omega_2(t) & \omega_3(t) \end{bmatrix}^T$ where

$$\omega_1(t) = \begin{cases} 0.2\sin^2\left(3(t-0.2)\right), & t \geq 0.2 \text{ second} \\ 0, & \text{else} \end{cases}$$

$$\omega_2(t) = \begin{cases} 0.3\sin^2\left(4(t-0.3)\right), & t \geq 0.3 \text{ second} \\ 0, & \text{else} \end{cases}$$

Let the initial conditions be selected as $x(0)=(0.2,-0.3)^T$ and $\hat{x}(0)=(0,0,0,0)^T$. The result of simulation is shown in Figs. 1-2. It can be observed from Figure 1, the state estimate $\hat{x}(t)$ of the proposed observer (34) indeed converges asymptotically to the original state $x(t)$ of the fuzzy system (3). On other hand, it can be seen from the Figure 2, the estimated $\hat{\omega}(t)$ indeed converges asymptotically to the original disturbance $\omega(t)$ of the fuzzy system (3). One can see that the estimation performance is desired in the disturbance and perturbed environment.

5. CONCLUSION

In this paper, the T-S fuzzy observer design for the TS fuzzy system with unknown output disturbance and bounded time-varying input delay has been presented. Two theorems corresponding to the fuzzy systems with equal output matrices and with different output matrices, respectively, have been derived to give the existence condition for the estimator design. With the aids of LMI tool, the estimator design procedure has been summarized. Simulation results have been shown that the estimator design is effective and successful.

6. ACKNOWDLEGMENT

This work was supported by National Science Council, Taiwan, under Grant NSC 98-2221-E-008-093-MY3.

7. REFERENCES

[1] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116– 132, Jan. 1985.

[2] G. Feng, S. G. Cao, N. W. Rees, and C. K. Chak, Design of fuzzy control systems with guaranteed stability, *Fuzzy Sets and Systems.*, vol. 85, pp. 1–10, 1997.

[3] X. D. Liu and Q. L. Zhang, New approaches to H∞ controller designs based on fuzzy observer for T–S fuzzy systems via LMI, *Automatica*, vol. 39, pp. 1571–1582, 2003.

[4] K. Tanaka, T. Ikeda, and H. Wang, A multiple Lyapunov function approach to stabilization of fuzzy control systems, *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 4, pp. 582–589, Aug. 2003.

[5] K. Tanaka and H. O. Wang, *Fuzzy Control Systems Design and Analysis. A Linear Matrix Inequality Approach* (New York, Wiley, 2001).

[6] Z. Yi and P. A. Heng, Stability of fuzzy systems with bounded uncertain delays, *IEEE Transactions on Fuzzy Systems,* vol. 10, no. 1, pp. 92–97, Feb. 2002.

[7] T. V. Dang, W. J. Wang, C. H. Huang, C. H. Sun and L. Luoh, Observer Synthesis for the T-S Fuzzy with Uncertaintyand Output Disturbance, *Journal of Intelligent Fuzzy and Systems,* vol 22, no. 4, pp. 173-183, 2011.

[8] Y.Y. Cao and P.M. Frank, Analysis and synthesis of nonlinear time-delay system via fuzzy control approach, *IEEE Transactions on Fuzzy Systems*, vol. 8, 2000, pp. 200–211.

[9] K. Gu, V. L. Kharitonov and J. Chen, *Stability of time-delay systems* (Birkhauser, 2003).

[10] M. N. A. Parlakci, Improved robust stability and stabilizability criteria for time-delay systems with nonlinear perturbations, *the Proceeding of 16th IFAC World Congress*, Prague, 2005, pp. 6-8.

[11] C. Pen, D. Yue, T. C Yang, and E. G. Tian, On delaydependent approach for robust stability and stabilization of T-S fuzzy systems with constant delay and uncertainties, *IEEE Transactions on Fuzzy Systems*, vol. 17, 2009, pp. 1143-1156.

[12] J. P. Richard, Time-delay systems: an overview of some recent advances and open problems, *Automatica*, vol. 39, 2003, pp. 1667-1694.

[13] K. Busawon and P. Kabore, Disturbance attenuation using proportional integral observers, *International Journal of Control,* vol. 74, 2001, pp. 618-627.

[14] Z. Gao and D. W. C. Ho, Proportional multiple-integral observer design for descriptor systems with measurement output disturbances, *the Proceedings of the Institution of Electrical Engineers-Control Theory and Applications*, vol. 151, 2004, pp. 279-288.

[15] Z. Gao, X. Shi, and S.X. Ding, Fuzzy state/disturbance observer design for T–S fuzzy systems with application to sensor fault estimation, *IEEE Transactions on Systems, Man, and Cybernetics,* part B, vol. 38, 2008, pp. 875-880.

[16] Z. Xu and X. Li, Control design based on state observer for nonlinear delay systems via an LMI approach, *International Journal of Information and Systems Sciences*, vol. 4, 2008, pp. 84-88.

[17] D. G. Luenberger, An introduction to observers, *IEEE Transactions on Automatic Control,* vol. 16, 1971, pp. 596-602. [18] H. H. Choi, LMI-based nonlinear fuzzy observer-controller design for uncertain MIMO nonlinear systems, *IEEE Transactions on Fuzzy Systems,* vol. 15, 2007, pp. 956-971.

[18] D. W. Gu and F. W. Poon, A robust state observer scheme, *IEEE Transactions on Automatic. Control,* vol. 46, 2001, pp. 1958-1963.

[19] A. Hamzaoui, N. Essounbouli, K. Benmahammed, and J. Zaytoon, State observer based robust adaptive fuzzy controller for nonlinear uncertain and perturbed systems, *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 34, 2004, pp. 942-950.

[20] C. S. Tseng, Model reference output feedback fuzzy tracking control design for nonlinear discrete-time systems with time-delay, *IEEE Transactions on Fuzzy Systems.,* vol. 14, 2006, pp. 58-70.

[21] C. H. Tseng and B. S. Chen, Robust fuzzy observer-based fuzzy control design for nonlinear discrete-time systems with persistent bounded disturbances, *IEEE Transactions on Fuzzy Systems,* vol. 17, 2009, pp. 711-723.

[22] Z. Xu and X. Li, Control design based on state observer for nonlinear delay systems via an LMI approach, *International Journal of Information and Systems Sciences*, vol. 4, 2008, pp. 84-88.

[23] J. Yoneyama and M. Nishikawa, Output stabilization of Takagi–Sugeno fuzzy systems, *Fuzzy Sets and Systems,* vol.111, 2000, pp. 253–266.

[24] S. Boyd, L. E. Ghaoui, E. Feron and V. Balakrishnan, *Linear matrix inequalities in system and control theory* (Society for Industrial and Applied Mathematics, Philadelphia, 1994).

[25] K. Tanaka and H. O. Wang, *Fuzzy control systems design and analysis: A linear matrix inequality approach* (Wiley, New York, 2001).

[26] J. H. Park and O. M. Kwon, Guaranteed cost control of time-delaychaotic systems, *Chaos, Soliton, Fractal*, vol. 27, pp. 1011–1018, 2006.

[27] Y. Wang, L. Xie, and C. E. De Souza, Robust control of a class of uncertain nonlinear systems, *Systems and Control Letters*, vol. 19, pp. 139–149, 1992.

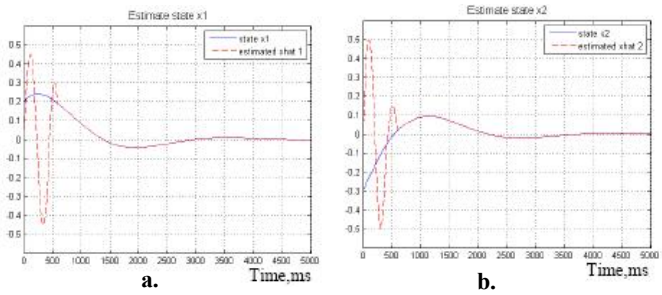


Figure 1. a, Estimate the state $x_1(t)$ b, Estimate the state $x_2(t)$

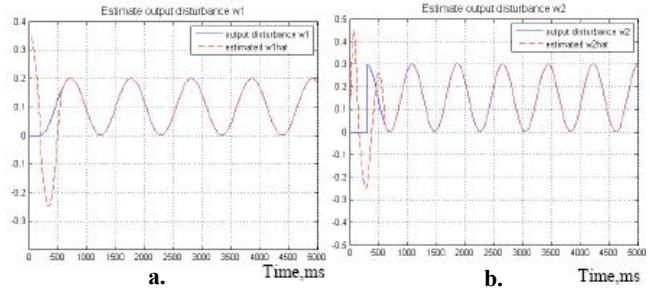


Figure 2. a, Estimate the state $\omega_1(t)$ b, Estimate the state $\omega_2(t)$

Diagnosis Dengue Fever and Typhoid Fever Using Fuzzy Logic Approach

Khodijah Hulliyah
State Islamic University (UIN) Jakarta
Tamansari Puribali Blok D1/24
Sawangan Depok
+628128947537
khodijah.hulliyah@uinjkt.ac.id

Siti Pratiningsih
State Islamic University (UIN) Jakarta
Jl.Ir.H. Juanda No.95
Ciputat Tangerang
+62856 9415 3223
pratiningsih.siti@gmail.com

ABSTRACT
Dengue hemorrhagic fever (DHF) is an endemic disease in Indonesia which, if accompanied by complications can cause death. One of the differential diagnosis of typhoid fever is a fever which also remains a serious problem in Indonesia. Dengue and typhoid fever has clinical manifestations similar symptoms, especially fever, this resulted in frequent errors early diagnosis of dengue and typhoid fever. Early diagnosis of typhoid fever, dengue fever and right will be very useful because it can avoid the occurrence of complications. To map the similarity of clinical symptoms of dengue and typhoid fever can then use fuzzy logic. This thesis aims to build a model with a fuzzy logic expert system approach for the clinical diagnosis of DHF and typhoid fever. System development method used is a RAD with Matlab as such tools. Test results show that the system can perform its function in the diagnosis of dengue and typhoid fever. Increase in the number of parameters and testing using a variety of fuzzy membership functions and various types as well as other methods of inference can be made for future development.

Keywords
fuzzy logic, diagnosis, dengue, typhoid fever, RAD

1. INTRODUCTION
Dengue hemorrhagic fever (DHF) is a disease caused by dengue virus I, II, III, and IV are transmitted mosquito Aedes aegypti and Aedes albopictus [16]. The disease is characterized by a sudden high fever accompanied by plasma leakage and bleeding, can cause death and cause outbreaks [4]. One of the major clinical manifestations in dengue are fever [16]. The pattern of fever with accompanying clinical symptoms is very important to know (http://www.pdpersi.co.id). In the early course of the disease one of the differential diagnosis of typhoid fever DHF is where both diseases are classified as tropical disease and is endemic in Indonesia. Typhoid fever in Indonesia is still an endemic disease that often cause problems and if accompanied complications can lead to death [14]. The success of efforts to address cases of DHF is primarily determined by the precision in the early diagnosis and management, including observation and treatment of blood pressure, pulse and prevention of fluid administration / overcome the shock [4]. Meanwhile, the early diagnosis of typhoid fever is very useful to immediately be given adequate treatment so as to avoid the onset of complications [14]. Complaints and symptoms of typhoid fever include fever, headache, dizziness, muscle pain, anorexia, nausea, vomiting, or diarrhea obstipasi [14]. Complaints and symptoms of dengue include fever, there is a manifestation of bleeding, headaches, sore muscles, bones and joints, nausea and vomiting [4] From the above description, it looks much similarity of clinical symptoms dengue and typhoid fever, although with different specific characteristics, so that errors may occur early diagnosis for patients and families of patients. This can lead to premature patient mishandling. Furthermore, if complications occur and causedeath.

In computer terminology the above problems can be termed as a complex mapping input space to output space. In this case the input space is the clinical symptoms of typhoid fever and dengue and output space corresponding to the type of disease with clinical symptoms of dengue and typhoid fever. Called complex because there are members of the input space DHF which also includes the input space into typhoid fever and vice versa. This problem can be solved by fuzzy logic.

Free maternity hospital (RBG) Alms Houses in East Jakarta as a means of free health care for citizens in need, contribute and help reduce the death rate due to dengue and typhoid fever in the community especially the ability of a weak economy. Among the services provided RBG besides maternal and child health services also include public health services, including dengue fever and typhoid fever. All services provided free of charge at the RBG. In addition to health services, RBG also provide guidance for maintaining cleanliness and health of one of them to know and distinguish the symptoms of the disease. Limited human resources at RBG become its own problems, so the idea to develop applications that can help experts be required. In this case the application is focused on dengue fever and typhoid fever.

Based on the above description, writer take title "Application of Fuzzy Logic and Procedure for Diagnosis of Diseases of Dengue Hemorrhagic Fever and Typhoid Fever."

2. APPROACH
2.1 Fuzzy Logic
Fuzzy logic is an appropriate way to map an input space into an output space by using a membership function (membership function). Several types of membership functions (mf) is a binary sigmoid, gaussian, generalized-bell, trapezoidal, and triangular. In fuzzy logic there are several processes, namely the determination of a fuzzy set, the application of IF-THEN rules and fuzzy inference process. Fuzzy inference methods include

methods Tsukamoto, Mamdani and Sugeno order 0 or order 1 [10].

2.2 Expert System

Expert systems are systems that try to adopt human knowledge to computer, so that computers can solve problems as they are commonly performed by experts or specialists [10].

2.3 Dengue Hemorrhagic Fever (DHF)

Dengue hemorrhagic fever (DHF) are caused by dengue virus belonging to group B Arthropod Borne Virus (Arboviroses) is now known as the genus Flavivirus, family Flaviviridae and has 4 types namely serotypes: DEN-1, DEN-2, DEN-3 and DEN-4 . The four serotypes of dengue virus can be found in various regions in Indonesia. DEN-3 serotype is the dominant serotype and assumed many of which showed severe clinical manifestations [4].

Classical form of dengue is characterized by high fever, sudden 2-7 days, accompanied by a reddish face. Complaints such as anorexia, headache, sore muscles, bones, joints, nausea and vomiting are common. Some patients complain of pain with swallowing farings hiperemis found on inspection, but seldom found cough and cold. The most common form of bleeding is the tourniquet test (Rumple leede) positive, easy bruising and bleeding skin on intravenous injection or blood sampling used in [4].

2.4 Typhoid fever

Typhoid fever is a systemic infectious disease caused by the bacteria Salmonella typhi and Salmonella bacteria sometimes paratyphi. The disease is transmitted through food or water contaminated by bacteria S. typhi [14]. Among the signs and symptoms caused in patients with typhoid fever include: fever, dirty tongue, the middle white and red edges, anorexia, severe nausea to vomiting, obstipasi or diarrhea, headache, muscle aches, and abdominal pain caused by swelling liver and spleen [14].

2.5 Matlab

Matlab stands for Matrix Laboratory, is a high-level programming language which is devoted to the needs of technical computing, visualization and programming such as computational mathematics, data analysis, algorithm development, simulation and modeling and computation graphs (www.mathworks.com)

3. SYSTEMS DEVELOPMENT METHODS

System development methods to be used is a Rapid Application Development (RAD). The following stages of system development RAD according to Pressman (2005):

a. Communication
b. Planning
c. Modelling
d. Construction
e. Deployment

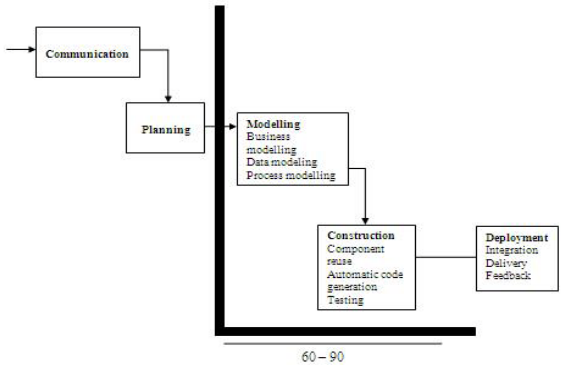


Figure 1. RAD Phases

4. RESULTS AND DISCUSSION

4.1 Communication

Based on the elaboration of the background which has been described previously as well as literature studies will require the development of fuzzy logic model for the diagnosis of dengue and typhoid fever, with expert systems approach to systems development that is expected to answer the problem.

4.2 Planning

Modeling of fuzzy logic will use MATLAB software for ease and reliability of computational considerations. The data used is the data of clinical symptoms of dengue fever and typhoid fever which have similarities with a different characteristic for each disease.

Working system interface application to adopt the workings of the expert system because it is the approach taken is an expert system approach. The application will accept input parameters of clinical symptoms and subsequent applications will provide the output of the diagnosis.

Parameters to be used is the clinical symptoms of dengue fever and typhoid fever include: fever, sore muscles and joints extrimis up (hands and arms) and extrimis bottom (foot), the manifestation of bleeding such as bleeding in the nose and gums and tornikuet test positive, the digestive disorders, as well as examination of the tongue are webbed or not.

4.3 Modelling

4.3.1 Business Modelling

The purpose of this application design is designing a fuzzy logic model for the diagnosis of dengue fever or typhoid fever based on the input of certain clinical symptoms The system requirements are the data of clinical symptoms of dengue fever and typhoid fever for the formation of fuzzy rules. While for the diagnosis of dengue fever and typhoid fever expert system approach is used. As for the interface program will be designed by using the tools Matlab 7.8.0. because of its superiority in numerical computation and data visualization.Subsubsections

4.3.2 Data Modelling

Design of data structures using a fuzzy set is divided into criteria and parameters. Criteria for the clinical symptoms of dengue

fever and typhoid fever include fever, muscle and joint pain, bleeding manifestations, digestive disorders and conditions of the tongue. Each criterion has a parameter that reflects the membership of a fuzzy set or membership function (mf). Criteria, parameters and the membership function value obtained by interviews with experts. For each criterion consists of three parameters with the same attributes but the attribute values of different linguistic. Attribute values ranging between 00-10. For low value of the attribute membership function (0 - 3.6) used a function of Z (zmf), attribute values are (1.69 - 5) used a Gaussian function (gaussmf) and attribute a high value (6.49 to 9.6) used functions of the form S (SMF).

Table 1. Fuzzy Value fever

Fever	Score	Measurement Value
Gradually	0,00 – 3,21	The body temperature rises gradually 1 week
Hesitation	1,69 – 5,00	in the morning intermittent fever and night
Suddenly	6,49 – 9,60	Appears suddenly, remain high for 2-3 days

Table 4.2 Value fuzzy sore muscles and joints

Muscles And Joints	Score	Measurement Value
Not interfere	0,00 – 3,60	No complaints
Disruptive	1,69 – 5,00	There have been complaints, but not too distracting
Highly interfere	6,49 – 9,60	is very disruptive and the patient complained

Table 4.3 Value fuzzy Bleeding

The bleeding manifestations	Score	Measurement Value
Unclear	0,00 – 3,60	Bleeding nose and gums a bit and not spontaneous
Clear	1,69 – 5,00	spontaneous bleeding nose and gums, test positive tornikuet
Highly clear	6,49 – 9,60	Haematemesis or melena

Table 4.4 Value fuzzy indigestion

Value indigestion	Score	Measurement Value
Occurred	0,00 – 3,60	constipation or diarrhea occurs with high frequency
Hesitation	1,69 – 5,00	Happen constipation or diarrhea but with low frequency
Not occurred	6,49 – 9,60	No constipation or diarrhea

Table 4.5 Value of fuzzy conditions of the tongue

The Tongue Condition	Score	Measurement Value
Webbed	0,00 – 3,60	Tongue dirty in the middle, and end of the red edge
Hesitation	1,69 – 5,00	is not clear whether or not webbed
No webbed	6,49 – 9,60	normal tongue color

Output is divided into 4 categories: diagnosis of typhoid fever, observations, laboratory checks and DHF.

Table 4.6 Value of output

Sample Data Type	Minimum Value	Maximum Value
Typhoid fever	0,00	3,99
Observation	3,00	5,99
Check the lab	5,00	7,99
Dengue Fever	7,00	10,00

4.4 Construction

4.4.1 Installing Matlab Program

Before going into the writing phase of the program, first conducted the installation program Matlab.

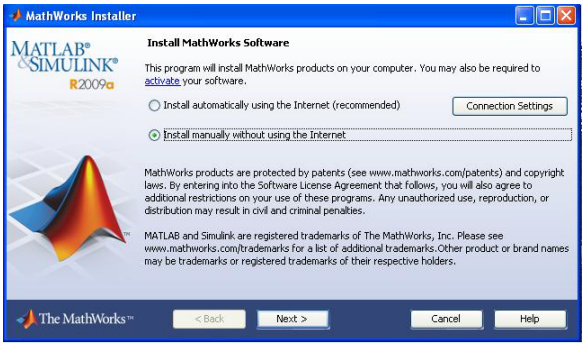


Figure 3. Input gejala 1

Matlab program has been successfully installed and the activation

Figure 3. Input gejala 1

4.4.2 Writing Program

At this stage, the authors translate the flowchart into a system of program codes. Writing program code to be used is by using Matlab 7.8.0

4.4.3 Testing Program

Tests performed using black box approach.

4.4.3.1 Unit Test

In this test, we will be to test for each module applications ranging from entry to exit from application to application. The following is the result of testing with black box approach

4.4.3.2 Integration test

Integration test performed to the overall testing program, beginning symptoms include symptoms of 1 to 5, window diagnosis until the result diagnosis.

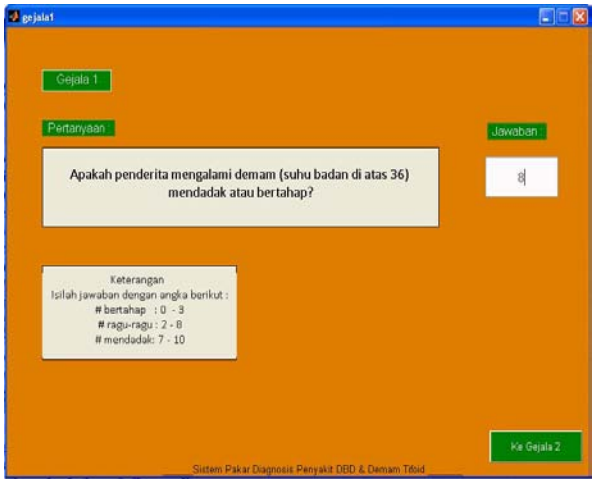


Figure 3. Input gejala 1

Figure 4. window diagnosis

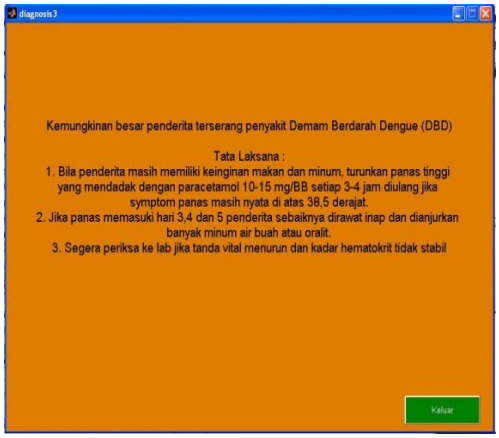


Figure 5. Result Diagnosis

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

Conclusions from the results of research conducted as follows:

a. Precise determination of fuzzy logic is used to map the input and output of complex diseases such as dengue fever and symptoms of typhoid fever.
b. Systems with expert systems approach can provide the diagnosis and management of DHF and typhoid fever.

5.2 Recommendations

For further development, the author gives suggestions:

a. Adding the number of input parameters of symptoms
b. Designing the input interface in the form of linguistic variables
c. Tested with a variety of input membership functions, type of membership functions and inference methods are different.
d. Making web-based system

REFERENCES

[1] Anonymous. 2008. Beware of Dengue Hemorrhagic Fever. [Online] Available: 3Akesehatan & Itemid = 66 & lang = en [July 16, 2010 14:00]

[2] Anonymous. 2009. Diseases Dengue Hemorrhagic Fever (DHF)-Definition, Causes and Symptoms of dengue. [Online] Available: http://organisasi.org/penyakit-demam-berdarah-dengue-dbd-pengertian-penyabab-gejala-dbd [July 16, 2010 14:00]

[3] Challoner, Jack. 2003. Artificial Intelligence, A Guide for Beginners to Robotics and Artificial Human Intellect. New York: Erlangga

[4] Ministry of Health Republic of Indonesia Directorate General of Communicable Disease and Environmental Health. 2004. Governance of Dengue Hemorrhagic Fever in Indonesia. The third edition. Editor of Sri Rezeki et al. London: Department of Health

[5] Directorate of Law and Information. Programming Basics. [Online] Available:

http://www.djkn.depkeu.go.id/download/modul/Dasar-Dasar_Pemrograman.pdf [October 12, 2009 11:30]

[6] Fitriyanti. 2010. Diagnosis of Pulmonary TB Disease and Bronchial Asthma Using Backpropagation Artificial Neural Network Method. Informatics Engineering Program Syarif Hidayatullah State Islamic University. Thesis Not Published

[7] Hartanto, Y. Thomas Rev & Rev Agung Dwi Prasetyo. 2003. Control System Analysis and Design with Matlab. London: ANDI

[8] Harvey, A. McGehee, et al. 1991. Diagnosis of Appeal (Case Oriented Clinic). New York: Literacy Binarupa

[9] Kendall, Kenneth E. and Julie E. Kendall. 2006. Analysis and

Design System Language Indonesia.Ed.5 Edition Volume 1. New York: INDEX

[10] Kusumadewi, Sri. 2003. Artificial Intelligence, Engineering and Application. London: Graha Science

[11] Naba, Agus. 2009. Quick Learning Fuzzy Logic Using MATLAB. London: ANDI

[12] Novriani, Harli. 2002. Immune Response and Degree of Morbidity and Dengue Hemorrhagic Fever Dengue Shock sydrome. PowerPoint Presentation (134), 46-47 [Online] Available: [July 16, 2010 13:00]

[13] Nugroho Adi. 2005. Information Systems Analysis and Design with Object Oriented Methodology. New York: Information

[14] Pressman, Roger S. 2005. Software Engineering, A Practitiner Approach. Sixth Edition. New York: McGraw Hill

[15] Proboyekti, Umi. Software Process Model I. [Online] Available: http://lecturer.ukdw.ac.id/othie/softwareprocess.pdf [October 13, 2009 08:30]

[16] Day, Jong jack. 2009. Artificial Neural Networks Using MATLAB & programming. London

[17] Soegijanto, Soegeng. Collection of Papers of Tropical and Infectious Diseases in Indonesia, Volume 1. 2004. New York: Airlangga University Press

[17] Scientific Center.2002.Kamus Popular Media Team. New York: Media Center

[18] Turban, Efraim. 1994. Decision Support and Expert Systems Management Support Systems.ed.4. New York: Prentice-Hall

Website:
http://www.mathworks.com/
http://www.pdpersi.co.id/
http://pusatbahasa.diknas.go.id/
http://kamusbahasaindonesia.org/
http://www.kamus-medis.co.cc/
http://digilib.its.ac.id/

Customer Satisfaction Control Application in Quality Assurance Departement at Petra Christian University using Fuzzy Aggregation

Andreas Handojo
Informatics Departement
Petra Christian University
Jl. Siwalankerto 121-131,
Surabaya
Indonesia
handojo@petra.ac.id

Rolly Intan
Informatics Departement
Petra Christian University
Jl. Siwalankerto 121-131,
Surabaya
Indonesia
rintan@petra.ac.id

Denny Gunawan
Informatics Departement
Petra Christian University
Jl. Siwalankerto 121-131,
Surabaya
Indonesia

ABSTRACT
As an institution that want to give a good service to their customer, Petra Christian University need to watch over the service quality that each department/unit give to their customer. To do it, Quality Assurance Department (QAD) was created to measure, supervise and maintain the customer satisfaction level in the university. To do that, QAD must distribute many questionnaires from each department/unit to their own customer which is include all university students, administration staff, and faculty member. And then, the result from that questionnaire must be calculated to produce customer satisfaction level, which is need a lot of time.

In this research, an Customer Satisfaction Control Application was created to help Quality Assurance Department to distribute questionnaire, and to help in measuring the performance of every department/unit in Petra Christian University. This application will be build using fuzzy aggregation to examine the customer satisfaction questionnaire result and build using MYSQL and PHP.

Based on testing, this application has helped the users in evaluating the performance, and with the usage of graphics which explain the result of the questionnaire in more detail, and distributing the questionnaires to most of university students, administration staff, and lecturers.

Keywords
Fuzzy Aggregation, Customer Satisfaction, Quality Assurance Departement.

1. INTRODUCTION
As an educational institution, Petra Christian University has an obligation to maintain their quality service that provided to each civitas academica and stakeholders. Based on that a Quality Assurance Department (QAD) has been established to measure, supervise, and maintain the university services quality, by accommodating all of comments and suggestions from students, administrative staff, and faculty member that collect every semester via questionnarie.

To achieve this objective, one of the activities carried by QAD is to distribute the questionnaire from each department/unit to their customer (that can be students, administration staff, or faculty member). The purpose of this questionnaire is to evaluate the satisfaction level from students, administration staff, and faculty member of the University toward all of the services that provided by the university. Using this questionnaire all criticisms, suggestions, and recommend about all the service can become an input to Petra Christian University to improving a better education.

The Difficulties that encountered by QAD is that the number of subjects (that must be cover) from the questionnaire is so big because the university has around 29 academic department and 15 non academic department/unit (such as Biro, library, and other service center) and to process this questionnaire it will be takes a lot of time. Besides that, there is another difficulty, how to assess customer satisfaction levels that have a qualitative input to a quantitative result.

Therefore, this research try to create an customer satisfaction control application that build base on website, so make the questionnaire from each department easy to distribute to their each customer and this application also process the questionnaire result to become level of customer satisfaction using fuzzy aggregation method to accomodate the qualitative problem.

2. FUZZY SET
Fuzzy set is a generalization of crisp sets in which each member of the set was characterized by a membership function. Range/domain values that used as a membership function values have a value between 0 and 1 [2]. This membership function of a fuzzy set A of a universal set X is expressed as μA:

$$\mu A : X \rightarrow [0,1] \quad (1)$$

For example the fuzzy set that describes the situation "approaching value 3" can be seen in Figure 1.

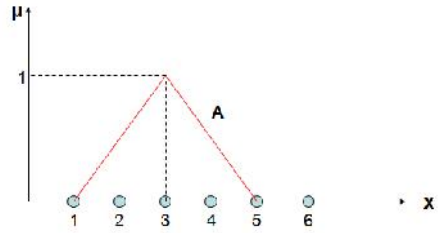


Figure 1. Fuzzy Set "approaching value 3"

The description of Figure 1 is as follows:
• A (3) = 1 and A (x) <1 for all x ≠ 3
• A symmetrical peak of x is worth 3 points
• A (x) decreases from 1 to 0 on each difference

$$\left|\frac{(3-x)}{2}\right| \quad (2)$$

Overall, for A can be expressed: [3]

$$A(x)=\begin{cases} 1-\left|\frac{(3-x)}{2}\right|, & for\ 1\le x\le 5 \\ 0, & for\ x<1\ and\ x>5 \end{cases} \quad (3)$$

2.1 Fuzzy Aggregation

Aggregation is a process of combining the values (numeric or non numeric) into one, so the result of aggregation is the result of the calculation of these values [2]. Aggregation Operators are divided into three general classes, and each class has a different nature or meaning, namely:

2.1.1 Conjunctive Operators

Conjunctive operators function is to combine multiple values by using the operator "and", so the result of the combination is high if and only if all the combined value of high [2]. Triangular norms (t-norms) are used in the conjunctive operators for fuzzy sets are:

$$f(x,y)\rightarrow \min(x,y) \quad (4)$$

2.1.2 Disjunctive Operators

Disjunctive operators function is to combine multiple values by using the operator "or", so the result of the combination is high if one or some combination of high value [2]. Triangular conforms (t-conorms) are used in disjunctive operators for fuzzy sets are:

$$f(x,y)\rightarrow \max(x,y) \quad (5)$$

2.1.3 Averaging Operators

Between conjunctive and disjunctive operators, there is one more class of operators which is averaging operators. The value of averaging operators is located in between the minimum and maximum value, which is the distance between t-norms and t-conorms [2]. Averaging operators have compensative properties (offset), where a low value could be offset by the high yield combined with intermediate value.

$$\min(x,y)\rightarrow f(x,y)\rightarrow \max(x,y) \quad (6)$$

2.1.4 Ordered Weighted Averaging (OWA) Operators

By using OWA operators [1], some fuzzy values can be combined as in the averaging operator by giving the weight to each condition, which $w=<w_1,w_2,...,w_n>$ is the OWA weight to calculate the value of a set of values $\{a_1,a_2,...,a_n\}$ can be generated by the following equation:

$$h_w(a_1,a_2,...,a_n)=w_1b_1+w_2b_2+...+w_nb_n \quad (7)$$

Where's $\{b_1,b_2,...,b_n\}$ is the permutation value from $\{a_1,a_2,...,a_n\}$ that create to make $b_1\ge b_2\ge b_3\ge...\ge b_{n-1}\ge b_n$. Example: assume that $w=[0.4,0.3,0.2,0.1]$ and $h_w(0.7,1.0,3.0,0.6)$. If it is sorted, then we will get b1 = 1.0, b2 = 0.7, b3 = 0.6, and b4 = 0.3. By doing aggregation, it will get:

hw(0.7, 1, 0.3, 0.6) = (0.4)(1) + (0.3)(0.7) + (0.2)(0.6) + (0.1)(0.3) = 0.76

3. DESIGN AND IMPLEMENTATION

Generally, the process from this system can be seen in Figure 2.

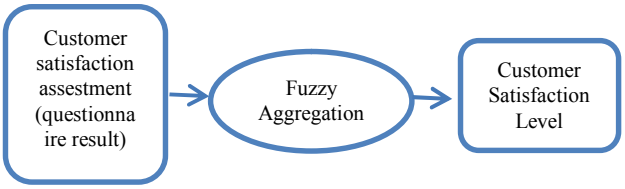


Figure 2. System Process Outline

At Petra Christian University, customer satisfaction assessment process carried out by a standard assessment which covers five aspects of service such as reliability, tangible, responsiveness, assurance, and empathy. This all five aspects will be cover on each questionnaire that has been distributed to all university members (student, administration staff, and faculty member) via website. The result from this questionnaire will automatically calculate on the website, which is the level of importance and the level of satisfaction about the service that has been given by each department/unit to their customer, and the difference between that two will become their performance result.

3.1 Calculating The Performance Using Fuzzy Aggregation

To calculate service quality performance that has been given by each department/unit. First, we define the fuzzy membership level of the importance (Figure 3), level of satisfaction (Figure 4), and

performance result (Figure 5), along with performance rules that are used in the calculation process (Table 1).

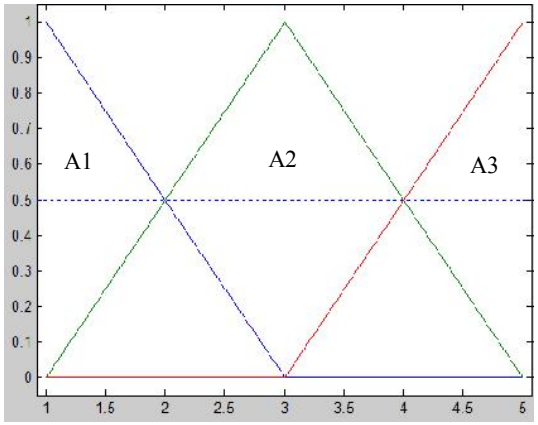


Figure 3. Fuzzy Membership for Level of The Importance

A1 = Not Important
A2 = Quite Important
A3 = Important

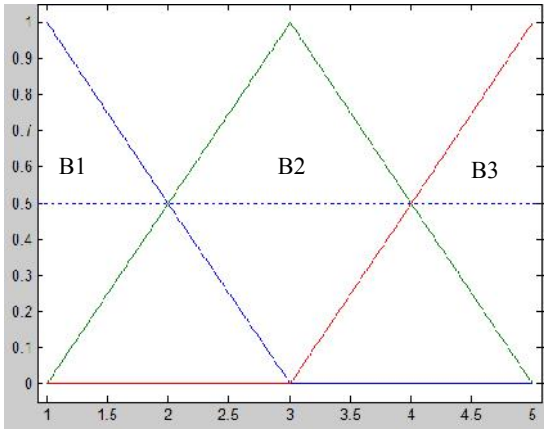


Figure 4. Fuzzy Membership for Level of Satisfaction Value

B1 = Not Satisfied

B2 = Quite Satisfied
B3 = Satisfied

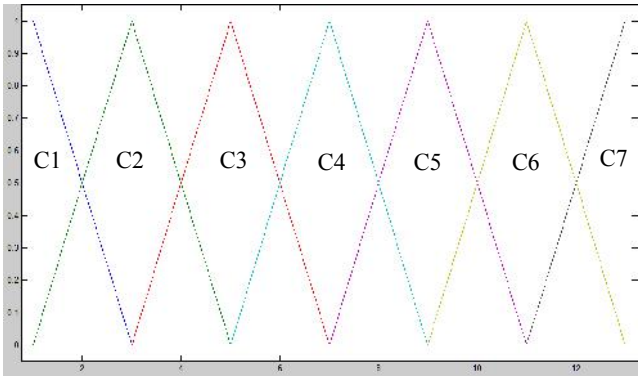


Figure 5. Fuzzy Membership for Performance Result

C1 = Very Poor
C2 = Poor
C3 = Fair Enough
C4 = Fair
C5 = Good Enough
C6 = Good
C7 = Very Good

On Table 1 show an example of rules that used to calculate the performance result using Fuzzy.

Table 1. Performance Rules that Used in Calculating Using Fuzzy

Level of Importance	&	Level of Satisfaction	→	Performance Result
A1	&	B1	→	C1
A2	&	B1	→	C2
A3	&	B1	→	C3
A1	&	B2	→	C4
A2	&	B2	→	C5
A3	&	B2	→	C5
A1	&	B3	→	C4
A2	&	B3	→	C6
A3	&	B3	→	C7

The next step is to get the membership of the input value of the level of importance and satisfaction for each of the existing rules. Then for any rules, find the minimum value and then get the maximum value from the minimum values that obtained before. After that, we will get the performance results according to table the performance results, using the equation below:

$$Performance\ result = \frac{\sum (wide\ x\ weight)}{\sum Wide} \quad (3.1)$$

After that, by using ordered weighted averaging operators [5] we can get the value of the performance by calculating the predetermined weights and the average point.

On the application, before the questionnaire distribute to the respondent, the administrator from QAD first must determined the fuzzy membership value from level of the importance, level of satisfaction, and performance result, along with rules that are used in the calculation process as show in Figure 6.

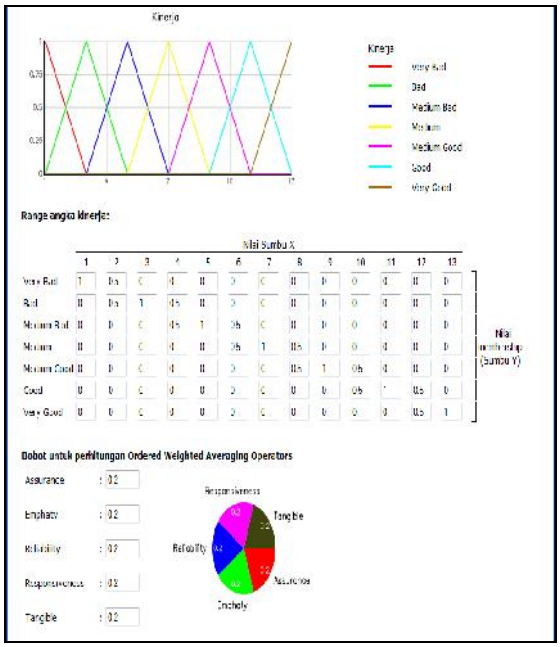


Figure 6. Determined the Fuzzy Membership Value

Then, the administrator can create the question from the questionnaire that wants to distribute (Figure 7). This questionnaire will save as a question data bank that can be use to build a questionnaire template (Figure 8).

This question can be group into the five aspects like tangible, reliability, responsiveness, assurance, and empathy as standard assessment.

So the template can be use as a master from the questionnaire that can be use in many department/unit, especially department/unit that give similar services or have same target respondent. This template also can be edited according specialty from each department/unit.

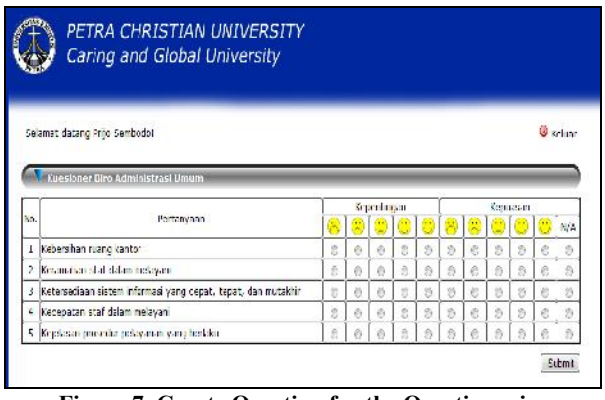


Figure 7. Create Question for the Questionnaire

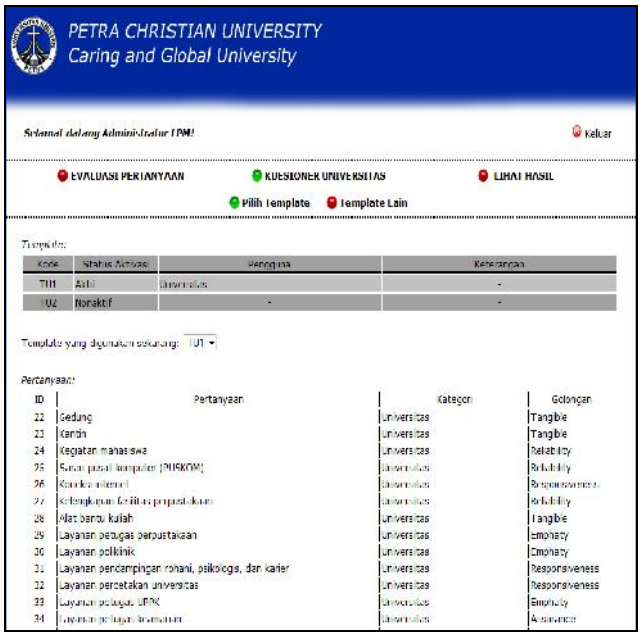


Figure 8. Manage Questionnaire Template to Distribute

After the questionnaires has been distributed and filled out by the respondents, then after a fixed time, this distribution process has been closed and the calculation performance process has been performed. The performance result can be show on bar chart (Figure 9) or as a fuzzy membership (Figure 10).

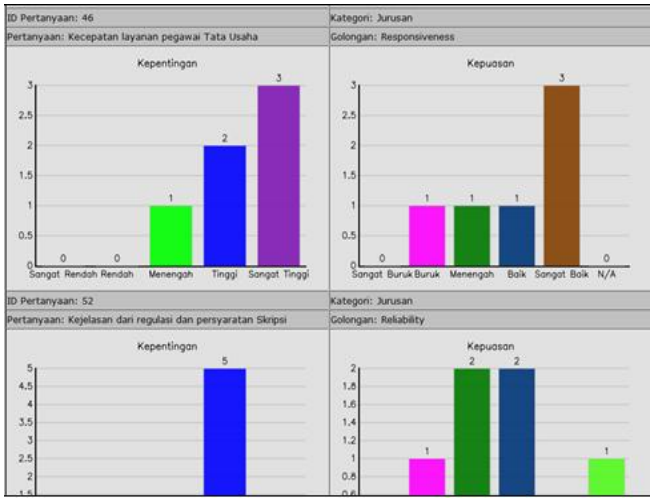


Figure 9. Performance Result on Bar Chart

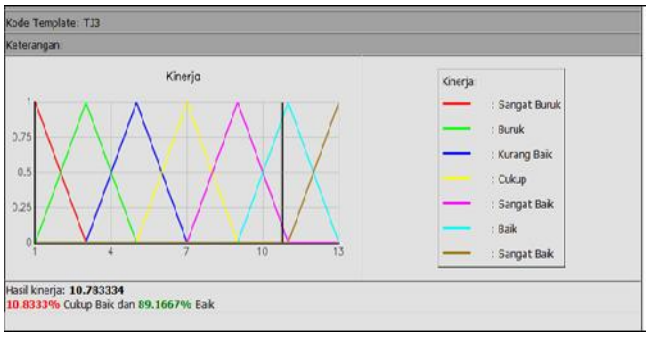


Figure 10. Performance Result on Fuzzy Membership

4. CONCLUSION

This application has helped the QAD to distribute the questionnaire from many department/unit on university to their customer to know their service performance on give the best service to their customer.

By applying fuzzy aggregation for determine level of the importance, satisfaction, and performance result, along with performance rules that are used in the calculation process, can solve the problem on evaluating the performance, and with the usage of graphics which explain the result of the questionnaire in more detail view.

5. REFERENCES

[1] F.ChicIana, F.Herrera, E.Herrera-Viedma and L.Martinez, 2003. A Note on the Reciprocity in the Aggregation of Fuzzy Preference Relations Using OWA Operators, Fuzzy Sets and Systems.

[2] Grabisch, M., Orlovski, S.A., and Yager, R.R. 1999. *Fuzzy aggregation of numerical preferences*. Norwell, United States of America: Kluwer Academic Publisher

[3] Hellman, M. 2003. *Fuzzy logic introduction*. France: Laboratoire Antennes Radar Telecom

[4] Klir, G., Yuan, B. 1995. *Fuzzy sets and fuzzy logic: Theory and applications*. Upper Saddle River, United States of America: Prentice Hall Inc.

[5] Peneva, V and Popchev, I. 2007. Aggregation of Fuzzy Relations with Fuzzy Weighted Coefficients.

A Comparison of Rabin Karp and Semantic-Based Plagiarism Detection

Catur Supriyanto
Department of Computer Science
Dian Nuswantoro University
Semarang, Indonesia
caturs@research.dinus.ac.id

Sindhu Rakasiwi
Department of Computer Science
Dian Nuswantoro University
Semarang, Indonesia
ochin_sindhu@yahoo.com

Abdul Syukur
Department of Computer Science
Dian Nuswantoro University
Semarang, Indonesia
abdul.syukur@research.dinus.ac.id

ABSTRACT
Document plagiarism is a challenging task for scholars. Similarity computation of two documents is the main step of document plagiarism. The accuracy of Rabin Karp and semantic-based document plagiarism is measured for comparison. This paper employed Latent Semantic Analysis (LSA) approach via Singular Value Decomposition (SVD) as the semantic-based document plagiarism. The result showed Rabin Karp has better performance than LSA Plagiarism.

Keywords
Plagiarism Detection, Rabin Karp, Latent Semantic Analysis.

1. INTRODUCTION
Plagiarism is the use of ideas or writing without enclosing the source of text. Students or lecturers who do research are very easy to do plagiarism [1]. It can be prevented by using a plagiarism detector to detect a plagiarism in a digital document.
According to [1], plagiarism detection method can be classified into several types. Based on complexity of the use method, plagiarism can be classified into superficial and structural. There is no linguistic rule in superficial type, different from structural type which used linguistic rule. Based on the number of document, plagiarism is classified into four categories, singular, paired, multidimensional and corpal. Singular plagiarism used single-document to compute the metric. Paired plagiarism used two document to be processed together to compute the metric. Multidimensional plagiarism used multi-documents to be processed together to compute the metric. Corpal plagiarism used all documents in the dataset to be processed together to compute the metric.

Based on the existing of reference (original) document, plagiarism can be split into two categories: external and internal plagiarism [2]. The difference is external plagiarism use reference document to detect plagiarism in suspicious document, meanwhile internal plagiarism identify the plagiarism in suspicious document without the existing of reference document.
Another type of plagiarisms is semantic-based and string matching-based plagiarism. Semantic-based plagiarism used transformation matrix to find the semantic relationship between terms in the corpus. The popular matrix transformation is Singular Value Decomposition (SVD). Meanwhile, string matching-based is string searching algorithm that can be used to plagiarism detection [3]. Detailed classification of document plagiarism is presented in [4].

The outline of this paper is as follows: section 2 describes the summary of algorithms. Section 3 describes data corpus. Section 4 shows the performance analysis of two algorithms. Section 5 presents the conclusion and future work.

2. SUMMARY OF ALGORITHM
Two different plagiarism document approaches used in this paper are described in this section.

2.1 Rabin Karp Algorithm
Rabin Karp Algorithm is a searching method by using hash function. The purpose of hash function is to speed up the search. Rabin Karp has been implemented for plagiarism purpose, since it is impractical method to detect a plagiarized document. Rabin Karp algorithm can be seen as follow:

```
RabinKarpMatcher (String P, String T, integer d, integer q )
  n : length[T], m: length[P]
  h : d^(m-1) mod q
  p : 0; t_n : 0
  for  i=1 to m do
      p=d p+P[i] mod q
      t_n = d t_n +T[i] mod q
  for s=0 to n-1 do
     if p= t_n then
        if p[1...m]=T[s+1 .... S+m]  then
            print s
     if s<n-1 then
        t_n = (d {t_n – T[s+1]  h}+ T[s+m+1]mod q
```

Figure 1. Pseudo code of Rabin Karp Algorithm [5]

2.2 Latent Semantic Analysis
Latent Semantic Analysis (LSA) via Singular Value Decomposition (SVD) is a popular transformation technique in information retrieval area, such as document clustering, document summarization and document plagiarism [6][7][8]. SVD can reduce large dimensional space into lower dimensional space [9]. Moreover, SVD can enable capture semantic relationship between terms and the context. Therefore, the performance of information retrieval technique can be increased by using lower dimensional space or small number of terms.

The process of SVD starts with the creation of matrix a term by document matrix A. Matrix A has $m \times n$ dimensional space,

which m is the number of terms and n is the number of documents. The SVD of matrix A is defined as:

$$A_{m\times n} = U_{m\times k}\,\Sigma_{k\times k}\,V^{T}{}_{k\times n} \quad (1)$$

Where U is called left singular vector matrix, Σ is called singular value matrix and V^T is called right singular vector matrix.

Based on the above discussion, this paper used matrix V^T to compute the similarity of document, since V^T contains the vector of document.

3. DATA CORPUS

This paper used data corpus of plagiarized short answer developed by [10]. Data corpus1 consists of 100 documents (19 examples of each of the heavy revision, light revision and near copy levels and 38 non-plagiarized examples written independently from the Wikipedia source). For performances measure, we differentiated the corpus only into 2 categories, plagiarized and non-plagiarized document. Tokenization, stopword removal and stemming algorithm (porter stemming algorithm) as the preprocessing of document were implemented to the corpus in both algorithms.

Since, Rabin Karp and LSA-based plagiarism have difference approach to detect plagiarism in suspicious document, this paper implemented different similarity measure for both. We implemented dice similarity and cosines similarity for Rabin Karp and LSA-based document plagiarism respectively. The calculation of dice and cosines similarity is given bellow.

$$Dice(d_A, d_B) = \frac{2\left|w(d_A)\cap w(d_B)\right|}{\left|w(d_A)\right| + \left|w(d_B)\right|} \quad (2)$$

$$Cosines(d_A, d_B) = \frac{\sum w_A \times w_B}{\sqrt{\Sigma(w_A)^2}\times\sqrt{\Sigma(w_B)^2}} \quad (3)$$

Where $w(d_A)$ and $w(d_B)$ is word in document A and document B, w_A and w_B is the *tfidf* value of each term in document A and document B.

4. EXPERIMENT RESULT

Performance analysis of the algorithms is evaluated on the corpus collected for this paper. For the performance analysis, we choose an intrinsic evaluation method and used precision (P), recall (R), and F-measure (F). Similarity of two documents has a value in range from 0 to 1. 1 means that the documents are exactly the same and 0 means that documents are exactly different. A document is decided as plagiarized document if similarity of suspicious and original document is more than a threshold τ. This paper used the threshold τ between 0 % and 100%. The performance of both algorithms is performed in different n-grams (n=2, 3, 4). According to [11], the using of n-gram can identify the writer's style and n-gram gives some flexibility to detection task for the external plagiarism detection.

Table 1. A confusion matrix for two class imbalanced problem

System Predicted	Actual	
	Plagiarized	Non-Plagiarized
Plagiarized	True Positive (TP)	False Positive (FP)
Non-Plagiarized	False Negative (FN)	True Negative (TN)

By using a confusion matrix above, recall (R), precision (P) and F-measure (F) can be computed as follow:

$$R = TP / (TP + FN) \quad (4)$$

$$P = TP / (TP + FP) \quad (5)$$

$$F = \frac{2\times P\times R}{P + R} \quad (6)$$

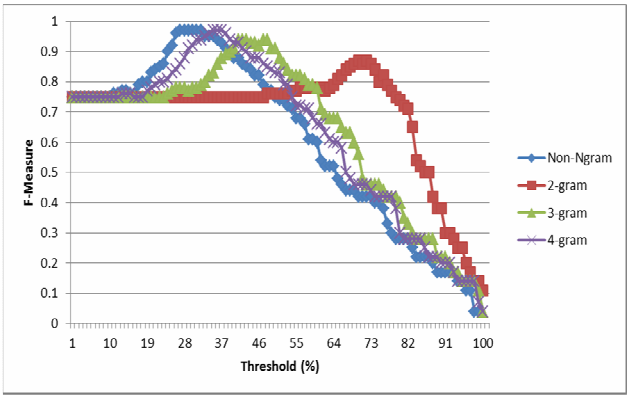


Figure 2. F-Measure of Rabin Karp Plagiarism Detection

Table 2 shows the performance evaluation result of Rabin Karp detection on the data corpus. The best F-measure 0.97 is obtained when the plagiarism does not use *n*-gram. For obtain the best result for each *n*-gram, threshold τ for 4-gram, 3-gram, 2-gram have been set into 35%, 47%, 71%, respectively. The smaller the number of *n*-gram, the higher the number of threshold τ is required.

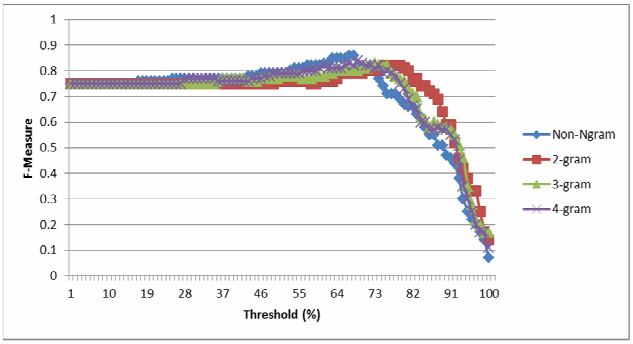


Figure 3. F-Measure of SVD Plagiarism Detection

[1] http://ir.shef.ac.uk/cloughie/resources/plagiarism_corpus.html

Table 3 shows the performance evaluation result of LSA based detection on the data corpus. The best F-measure 0.86 is obtained when the plagiarism also does not use *n*-gram. In LSA based plagiarism detection, the threshold τ has to be set more than 50% to obtain the best performance for non *n*-gram and each *n*-gram.

In comparison, from Figure 2 and Figure 3, experimental result shows that Rabin Karp performed better than LSA based plagiarism. Since, Rabin Karp use string matching approach which found the similar text directly. Different from LSA based plagiarism that needs to consider the noise in document collection (term-document matrix).

5. CONCLUSION AND FUTURE WORK

This paper compared two different plagiarism approaches between Rabin Karp and LSA based plagiarism. Although Rabin Karp plagiarism detection is simpler than LSA based plagiarism detection in detecting plagiarism in a document, the performance of Rabin Karp outperformed LSA based in plagiarism detection.

As a future research, we plan to evaluate other similarity measure to these approaches and try larger dataset to evaluate the performances. Also, machine learning can be used to plagiarism detection. The objective of machine learning in plagiarism detection is its ability to differentiate between original document and suspicious document automatically.

6. ACKNOWLEDGMENTS

The authors would like to thank Dian Nuswantoro University (UDINUS) for supporting this research.

7. REFERENCES

[1] Kashkur, M. and Parshutin, S. 2010. Research into Plagiarism Cases and Plagiarism Detection Methods. *Scientific Journal of Riga Technical University*. pp. 138-143.

[2] Zechner, M., Muhr, M., Kern, R., and Graz, K. 2009. External and Intrinsic Plagiarism Detection Using Vector Space Models. *In SEPLN 2009*.

[3] Singla, N. and Garg, D. 2012. String Matching Algorithms and their Applicability in various Applications. *International Journal of Soft Computing and Engineering (IJSCE)*. pp. 218-222.

[4] S. L, Thomas, B.B., and Idicula, S.M. 2011. A Study of Plagiarism Detection Tools and Technologies. *International Journal of Advanced Research in Technology*. vol. 1. pp. 64-70.

[5] Gupta, P., Agarwal, V. and Varshney, M. 2008. Design and Analysis of Algorithm, Asoke K. Ghosh, PHI Learning Private Limited.

[6] Abdulla, H.D., Polovincak, M. and Snasel, V. 2009. Using a Matrix Decomposition for Clustering Data. *International Conference on Computational Aspects of Social Network.*

[7] Gupta, V., Science, C. and Lehal, G.S. 2010. A Survey of Text Summarization Extractive Techniques. *Journal of Emerging Technologies in Web Intelligence*. vol. 2. pp. 258-268.

[8] Mudhasir, Y.S. 2011. Near-Duplicates Detection and Elimination Based on Web Provenance for Effective Web Search. In *Proceedings of the IJIDCS International Journal on Internet and Distributed Computing Systems*. pp. 22-32.

[9] Abidin, T.F. Yusuf, B. and Umran, M. 2010. Singular Value Decomposition for Dimensionality Reduction in Unsupervised Text Learning Problems. In *Proceedings of the ICETC International Conference on Education Technology and Computer*. pp. 422-426.

[10] Clough, P. and Stevenson, M. 2009. Developing A Corpus of Plagiarised Short Answers, Language Resources and Evaluation: Special Issue on Plagiarism and Authorship Analysis, In *Press. Journal Language Resources and Evaluation.*

[11] Stamatatos, E. 2009. Intrinsic Plagiarism Detection Using Character n-gram Profiles. In: Stein, B., Rosso, P., Stamatatos, E., Koppel, M., Agirre, E. (eds.) *SEPLN 2009 Workshop on Uncovering Plagiarism, Authorship, and Social Software Misuse (PAN 09)*. pp. 38-46.

Firefly Algorithm for Static Task Scheduling Problem

R. Eswari
Department of Computer Applications
National Institute of Technology Tiruchirappalli-620015
Tamilnadu, India
+91-431-2503744
eswari@nitt.edu

Nickolas Savarimuthu
Department of Computer Applications
National Institute of Technology Tiruchirappalli-620015
Tamilnadu, India
+91-431-2503739
nickolas@nitt.edu

ABSTRACT
Effective resource utilization can be achieved through proper scheduling of application tasks, which helps to attain better performance in the heterogeneous environment. Firefly algorithm, an efficient meta-heuristic algorithm is used in this paper to solve the task scheduling problem. This approach aims to generate optimal task schedule so as to get minimum completion time while executing application tasks. The efficiency of the algorithm with respect to makespan is compared with the existing particle swarm intelligence algorithm. The experimental results show that the firefly algorithm based approach outperforms PSO algorithm.

Keywords
Firefly algorithm, Heterogeneous computing, Particle swarm optimization, Task scheduling.

1. INTRODUCTION
A heterogeneous distributed computing system consists of machines with different computing capabilities and different computing speeds. To exploit the performance among these systems, scheduling of application tasks is to be considered as an important issue. The objective function of scheduling is to map the tasks onto the available processors and order their execution so that task precedence requirements are satisfied and minimum schedule length (or makespan) is obtained [1]. Generally, task scheduling algorithms are broadly classified into classes: static and dynamic. In static, the application characteristics such as execution times of tasks and data dependencies between tasks are known in advance, whereas in the dynamic scheduling decisions are made at run time. The static task scheduling for a heterogeneous distributed computing system is an NP-complete problem [1], which means that there is no known algorithm that finds the optimal solution in polynomial time.

Conference'04, Month 1–2, 2004, City, State, Country.

Several heuristic algorithms are proposed to solve the task scheduling problem, such as Simulated Annealing, Tabu Search, Swarm Intelligence which consists of Particle Swarm Optimization (PSO) and Ant Colony Optimization Algorithm (ACO). Xin-She Yang [2] developed Firefly Algorithm (FA) and used to solve continuous optimization problem. Simulation results of Xin-She Yang show that FA is superior to both PSO and GA in terms of efficiency and success rate. Lukasik and Zak also studied FA for continuous constrained optimization task. Their experiment demonstrated the efficiency of FA [3]. Because of these facts, in this paper firefly algorithm is considered to solve task scheduling problem.

The remainder of this paper is organized as follows. Section 2 describes the task scheduling problem. Section 3 gives an overview of the existing PSO method. Section 4 presents the Firefly Algorithm. Section 5 explains method of solving task scheduling based on firefly algorithm. Section 6 shows the effectiveness of the firefly algorithm and with experimental results and Section 7 concludes with future work.

2. PROBLEM DESCRIPTION
In a distributed environment, an application is decomposed into multiple tasks with data dependencies among them. It can be represented by a Directed Acyclic Graph (DAG), *G (T, E),* where *T* is the set of '*n'* tasks and *E* is the set of '*e'* edges between the tasks. Each task $t_i \in T$ represents a task in the distributed application, and each edge $(t_i, t_j) \in E$ represents a precedence constraint, such that the execution t_j starts after the execution of t_i. A task without any parent is called an entry task (t_{entry}), and a task without any child is called an exit task (t_{exit}). A sample DAG with communication costs and computation costs is given in Figure 1 and Table 1, respectively. Each edge $(t_i, t_j) \in E$ has a value that represents the communication overhead when data is transferred from task t_i to task t_j. A task can start execution on a processor only when all data from its parents become available to that processor.

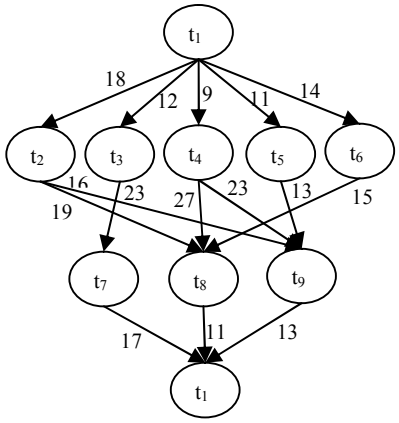


Figure 1. Sample DAG

The heterogeneous environment consists of a set of *m* processors connected in a fully connected topology in which all inter-processor communications are assumed to be performed without contention and computation can be overlapped with

communication. Task execution of a given application is assumed to be non-preemptive.

Table 1. Computation cost matrix

Task	M_0	M_1	M_2
t_1	14	16	9
t_2	13	19	18
t_3	11	13	19
t_4	13	8	17
t_5	12	13	10
t_6	13	16	9
t_7	7	15	11
t_8	5	11	14
t_9	18	12	20
t_{10}	21	7	10

The average computation cost of a task t_i is computed by the following equation.

$$\overline{w_i} = \sum_{j=1}^{m} w_{i,j}/m \quad (1)$$

where $w_{i,j}$ is the estimated execution time to complete t_i on processor m_j. The communication cost of the edge $e_{i,j}$ represents the cost of transferring $\mu_{i,j}$ units of data from task t_i scheduled on processor p_m to task t_j scheduled on processor p_n, is defined as

$$c_{i,j} = S_m + R_{m,n} * \mu_{i,j} \quad (2)$$

where S_m is the communication startup cost of p_m, $\mu_{i,j}$ is the amount of data transferred from task t_i to task t_j, and $R_{m,n}$ is the communication cost per transferred unit from p_m to p_n. The average communication cost of sending data from task t_i to task t_j is defined by

$$\overline{c_{i,j}} = \overline{S} + \overline{R} * \mu_{i,j} \quad (3)$$

where $\overline{S}$ is the average communication startup costs over all processors, $\overline{R}$ is the average communication cost per transferred unit over all processors. A task sequence for any application is found based on these average communication and computation costs. Tasks are selected one by one from the task sequence for execution and are assigned to a processor in the set of processors that minimizes its finish execution time using the insertion-based scheduling policy [1]. When a processor m_j is assigned a task t_i, the insertion based scheduling policy considers the possible insertion of the task in an earliest idle time slot between two already-scheduled tasks on that processor. This must be done without violating the precedence constraints among tasks. An idle time slot on processor m_j is defined as the difference between execution start time and finish time of two tasks that were consecutively scheduled on the processor m_j and should be greater than or equal to the computation cost of the task to be scheduled. The search starts from a time equal to the ready time of t_i on m_j, and proceeds until it finds the first idle time slot with a sufficient large time space to accommodate the computation cost of t_i on m_j. If no such idle time slot is found, the insertion-based scheduling policy inserts the selected task after the last scheduled task on m_j.

If two tasks t_i and t_j are on the same processor then $c_{i,j} = 0$ since intraprocessor communication is negligible. The Execution Start Time (EST) of task t_i on processor m_j is calculated by the equation (4) and (5).

$$EST(t_{entry}, m_j) = 0 \quad (4)$$

$$EST(t_i, m_j) = \max(avail[j], \max_{t_k \in pred(t_i)}(AFT(t_k) + c_{k,i})) \quad (5)$$

where $AFT(t_k)$ is the Actual Finish Time of a task t_k on the processor m_j, $avail[j]$ is the time that the processor m_j is free and it is ready to execute task t_i. The inner max block in equation (5) returns the ready time, i.e., the time when all data needed by t_i, has arrived at processor m_j.

To compute Execution Finish Time (EFT) of a task t_i, all immediate predecessor tasks of t_i must have been scheduled.

$$EFT(t_i, m_j) = w_{i,j} + EST(t_i, m_j) \quad (6)$$

If two processors are producing same EFT for a selected task, then the processor can be selected randomly or a lightly loaded processor can be selected.

After all tasks in a graph are scheduled, the schedule length (overall execution time) will be the AFT of the exit task t_{exit}. The schedule length also called makespan [1] is defined as

$$makespan = \max(AFT(t_{exit})) \quad (7)$$

The objective function of task scheduling problem is to assign tasks onto the available processors in such a way to produce minimum schedule length.

3. NATURE INSPIRED META HEURISTICS

In this section, the existing particle swarm optimization algorithm and the newly introduced firefly algorithm are studied. These algorithms are discussed below.

3.1 Particle Swarm Optimization Algorithm

The particle swam optimization (PSO) is a population-based stochastic optimization algorithm and is influenced by the simulation of social behavior of bird flocks. It was proposed by Kennedy and Eberhart [4]. In this PSO system, a swarm of particles fly through the search space. Each particle represents a candidate solution to the optimization problem. A particle is represented by its position vector X and its velocity vector V. Considering D as dimensions, then the velocity and position for the i-th particle is represented as $V_i=(v_{i1},.....,v_{iD})$ and $X_i=(x_{i1},.....,x_{iD})$.

Initially all particles are located randomly in their solution space. The position of a particle is changed by the best position by itself (its own experience) or the position of the best particle in its neighborhood (experience of neighborhood particles). The velocity of a particle is updated for each of its dimension $j \in 1,...,N_d$, hence, V_{ij} represents the velocity of j-th element of i-th particle. Thus the velocity of i-th particle is updated through the following equation:

$$v_{ij}^{k+1} = w * v_{ij}^{k} + c_1 * rand_1 * [pbest - x_{ij}^{k}] + c_2 * rand_2 * [gbest - x_{ij}^{k}] \quad (8)$$

where c_1 and c_2 are acceleration constants; r_1 and r_2 are two random numbers in the range [0,1]; w is the inertia weight which serves as a memory of previous velocities and controls the impact of previous velocities; is the velocity of the i-th particle in the k-th iteration, which represents the distance to be traveled by this particle from its current position; *pbest* represents its best previous position, and *gbest* represents the best position among all particles in the position. Velocity update is clamped with a user defined maximum velocity, V_{max}, which is used to prevent particles from premature convergence.

The position of j-th element of i-th particle is updated using the following equation:

$$x_{ij}^{k+1} = x_{ij}^{k} + v_{ij}^{k+1} \quad (9)$$

represents the particle position in the k-th iteration. Thus velocity and position of all particles are updated according to equations (8) and (9). This process is repeated until certain number of iterations reached. Best solution is measured in each iteration, which reflects the optimal solution.

3.2 Firefly Algorithm

The firefly algorithm is one of the nature-inspired meta-heuristic algorithms [5] and most powerful algorithms for solving the optimization problems. It is based on the social behavior of the fireflies flying in the sky. Most of the fireflies are capable of producing rhythmic flashes which is used to attract mating partners, and to attract potential prey. In addition to that the flashing light may also be served as a warning mechanism. The rhythmic flash, the flashing rate and the amount of time form part of the signal system that brings both sexes together.

The firefly algorithm was based on the idealized behavior of the flashing characteristics of fireflies. These flashing characteristics are idealized as the following three rules.

1. All fireflies are unisex so that one firefly will be attracted to other fireflies regardless of their sex.
2. Attractiveness is proportional to firefly brightness. For any two flashing fireflies, the less bright one will move towards the brighter one and the more brightness means less the distance. If no one is brighter than a particular firefly, it will move randomly.
3. The brightness of a firefly is determined by the value of the objective function.

Based on the above three rules, the basic steps of the firefly algorithm is given in [2].

The fireflies' communication mechanisms and their synchronization techniques have been imitated and effectively used in wireless network design, and robotics application. Permutation based flow shop scheduling, and independent task scheduling problems are also solved using this firefly algorithm. The scheme of FA is illustrated by the following pseudo code (figure 2).

Firefly Algorithm

Objective function f(x), x=(x_1, x_2 ..... x_m)
Generate initial solution of fireflies x_i (i=1,2,..,n)
Define light absorption coefficient γ
Repeat
  For *i = 1 : n all n fireflies*
    For *j = 1 : n all n fireflies*
      if $f(x_i) > f(x_j)$ *then*
        Find distance and attractiveness
        Move firefly i towards j in n-dimension
      End if
    End for *j*
  End for *i*
Rank the fireflies and find the current best
Until the stopping condition is met

Figure 2. Firefly Algorithm.

In the next part, certain issues of the firefly algorithm are considered.

3.2.1 Attractiveness

In the firefly algorithm, the attractiveness function $\beta(r)$ can be any monotonically decreasing function with the following generalized form:

$$\beta(r) = \beta_0 e^{-\gamma r^m}, \quad (m \geq 1) \quad (10)$$

where r is the distance between two fireflies, β_0 is the attractiveness at $r=0$ and γ is fixed light absorption coefficient which characterizes the variation of attractiveness value of firefly.

3.2.2 Distance

The distance between any two fireflies i and j at x_i and x_j is the Cartesian distance,

$$r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (11)$$

where $x_{i,k}$ is the k-th component of the i-th firefly (x_i).

3.2.3 Movement

The movement of a firefly i is attracted to another brighter firefly j is determined by

$$x_i = x_i + \beta_0 e^{-\gamma r_{ij}^2}(x_j - x_i) + \alpha(rand - \tfrac{1}{2}) \quad (12)$$

where the second term is the attractiveness function and the third term is randomization in which the randomization parameter α and random number generator "*rand*" are considered.

4. SCHEDULING USING FIREFLY ALGORITHM

In this paper, static task scheduling problem with the objective of minimizing the makespan is solved using firefly algorithm. A DAG with associated parameters such as computation costs and communication costs and the number of processors are the inputs

of the scheduling problem. The average computation cost of all tasks is calculated using equation (1). The processor allocation strategy is followed to calculate the firefly solution as given in section 2.

Initial solution: Let m refers to the number of fireflies and x_i refers to the position i at which firefly x is located. Each firefly represents a candidate solution (objective function) of the scheduling problem. Initially all fireflies are placed randomly at their positions and the one which contains minimum objective function ($x_{i\min}$) is considered and their positions are noted.

Distance: $f(x_i)$ represents the objective function of i-th firefly. All fireflies are compared with each other. If the objective function of firefly x_i is larger than the firefly x_j, then x_i moves towards x_j. The distance between the two fireflies x_i and x_j is calculated using Cartesian distance as given in equation (11).

Attractiveness: Based on the distance, initial attractiveness value and light absorption coefficient, the attractiveness function β is calculated using equation (10).

Movement: The movement of a firefly x_i is calculated using the following equation.

$$x_{i,k} \leftarrow (1-\beta)x_{i,k} + \beta x_{j,k} + \alpha\left(Random() - \frac{1}{2}\right); \forall k=1,\ldots,n \quad (13)$$

where $x_{i,k}$ is the k-th task of i-th firefly. $Random()$ is a random number obtained from the uniform distribution ($U(0,1)$). For every movement the firefly position is changed. The firefly with minimum objective function moves randomly.

$$x_{i\min} \leftarrow x_{i\min} + \alpha\left(Random() - \frac{1}{2}\right); \quad \forall k=1,\ldots,n \quad (14)$$

The above process is repeated for the user-defined maximum iteration. At the end of the last iteration, the firefly, which contains minimum value, is considered as best firefly and its value is taken as objective value.

5. EXPERIMENTAL RESULTS

This section presents a performance comparison of the firefly based scheduling algorithm and particle swarm based scheduling algorithm. Two benchmark application graphs given in [1] namely, randomly generated application graphs and real life application graphs called Fast Fourier Transformation (FFT) graphs and the performance metric (equation 15) suggested in [6] are used for experimental analysis. The algorithms are implemented using C language in Intel core i7-620M processor. The criteria of performance considered is to minimize the application makespan, which is the time elapsed between the first application task starts its execution and the last task completes its execution. The percentage improvement of firefly algorithm for scheduling the tasks with makespan reduction is computed by the equation given below,

$$\varphi = \left(1 - \frac{\sum Makespan_{FA}}{\sum Makespan_{PSO}}\right)\% \quad (15)$$

All tests were conducted for $\beta_0 = 1$, $\alpha = 0.2$ and fixed $\gamma = 1$ with population size $m = 40$ and iteration number $l = 500$. For PSO, the following parameters are considered, particle size = 40, number of iterations = 500, $c_1 = c_2 = 2$, inertial weight from 0.9 to 0.2. Both the algorithms were executed to solve all the instance over 10 trials, and the best trial was taken as the objective function value obtained.

5.1 Randomly generated application graphs

Random graphs are generated with several input parameters such as number of tasks, shape parameter (α), communication to computation ratio (CCR), and heterogeneity factor (β). For experimental purpose, more than 1000 random graphs are generated, among these around 100 test instances are taken from the combination of the sets of processors (2, 4, 8, 16, 32), tasks (20, 40, 60, 80, 100), CCR (0.5, 1, 1.5, 2), and β = (0.1, 0.25, 0.5, 0.75, 1.0), α=0.5.

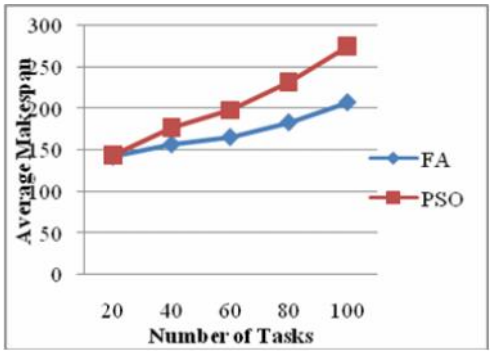


Figure 3. Average Makespan for different number of tasks

The average makespan obtained by firefly based scheduling algorithm with respect to various number of tasks, is compared with particle swarm based scheduling algorithm and the outcomes are shown in Figure 3. When the number of tasks is increased the completion time of the application obtained by the firefly algorithm is greatly reduced. The performance of the algorithms is evaluated with respect to different number of processors and CCR values. The firefly based algorithm reduces the completion time of an application even when limited numbers of resources are available as observed from Figure 4, and performs well for all CCR values as shown in Figure 5.

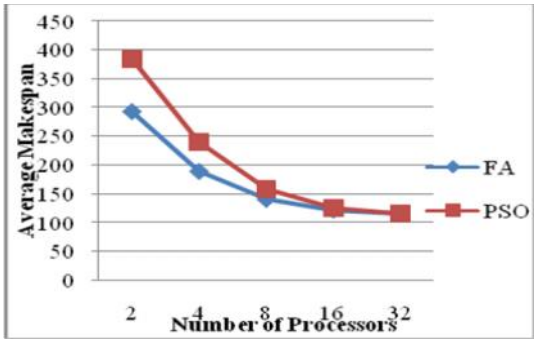


Figure 4. Average Makespan for different number of processors

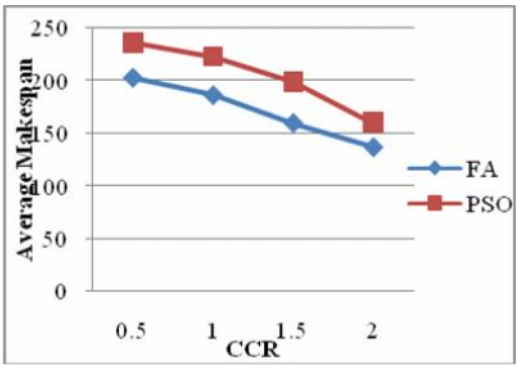


Figure 5. Average Makespan for different CCR

Finally, the performance improvement percentage of firefly algorithm for randomly generated graphs is calculated using equation (15). On the average firefly based solution quality is 16.1865% better than that of PSO and is shown in Table 2.

Table 2. Performance improvement percentage for random graphs

	Makespan	
Tasks	**FA**	**PSO**
20	142.35	142.95
40	157.15	176.4
60	165.85	197.45
80	183.65	230.9
100	207.7	274.45
Total	856.7	1022.15
Improvement of the algorithm (φ) = 16.1865%		

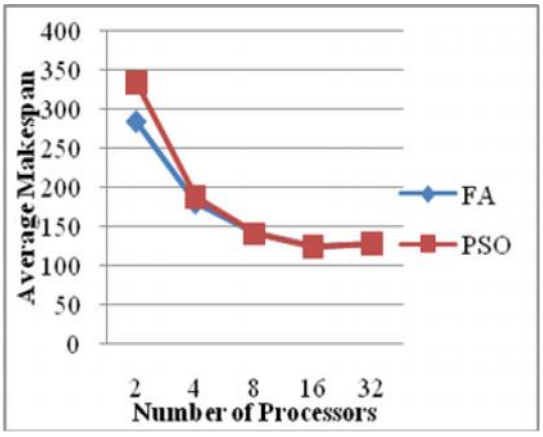


Figure 7. Average Makespan for different number of processors

5.2 Fast Fourier Transformation (FFT) application graphs

The application graph of a real world problem called FFT is considered. For this experiment, the graph characteristics such as CCR, β, α, and input vector M are considered. For an input vector of M, the number of nodes in the task graph is equal to $(2*M-1)+(M*\log_2 M)$ Since the structure of the application is known, other parameters are not needed. The number of processors considered is {2, 4, 8, 16, 32}. The number of data points in FFT varies from 4 to 16 incrementing powers of 2. Around 56 test instances are taken from the combination of the above parameters specified for this experiment.

The performance of the firefly algorithm is tested for different data points and is shown in Figure 6. The average makespan obtained for FFT graphs by the firefly algorithm with respect to different number of processors and different CCR values, is shown in Figure 7 and Figure 8 respectively.

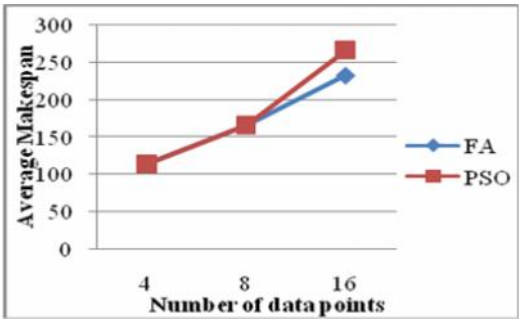


Figure 6. Average Makespan for different data points

The proposed scheduling algorithm outperforms when limited numbers of processors are available and gives minimum schedule length for all CCR values. Similar to results obtained for earlier random application graphs, the average makespan is decreased while increasing the number of tasks.

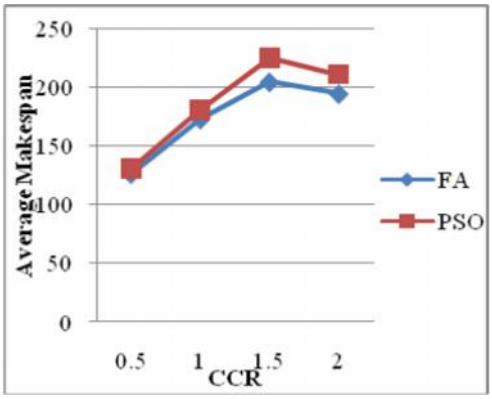


Figure 8. Average Makespan for different CCR

The performance improvement in makespan is also calculated for FFT graphs using equation (15) and is tabulated in Table 3. Firefly algorithm gives 6.3246% of improvement in makespan when compared with PSO algorithm.

Table 3. Performance improvement percentage for FFT graphs

	Makespan	
Tasks	**FA**	**PSO**
15	113.8125	113.9375
39	165.3	165.95
95	232.25	266
Total	511.3625	545.8875
Improvement of the algorithm (φ) = 6.3246%		

6. CONCLUSION

In this paper, static task scheduling problem is solved using firefly algorithm. The efficiency of firefly based scheduling algorithm is compared with existing PSO based scheduling algorithm. The experimental results show that the firefly algorithm outperforms PSO in most of the cases and produces minimum makespan while scheduling large number of tasks with less number of processors. Hence, efficient processor utilization is achieved through firefly based scheduling algorithm. The algorithm can be tested for more number of trials and for evaluating various numbers of real world applications as future course of research work.

7. REFERENCES

[1] Topcuoglu, H., Hariri, S., and Wu, M. Y. 2002. Performance-Effective and Low-Complexity Task Scheduling for Heterogeneous Computing. IEEE Trans. Parallel and Distributed Systems. 15, 5 (March 2002), 260-274.

[2] Yang, X. S. 2009. Firefly Algorithms for Multimodal optimization. In: Stochastic Algorithms: Foundations and Applications. SAGA, LNCS, 5792, 169-178. Springer, Heidelberg.

[3] Lukasik, S., and Zak, S. 2010. Firefly Algorithm for Continuous Constrained Optimization Tasks. ICCCI 2009, Lecture Notes in Artificial Intelligence (Eds. N. T. Ngugen, R. Kowalczyk. S. M. Chen), 5796, 97-100.

[4] Kennedy, J., Eberhart, R. C. 1995. Particle Swarm Optimization. In Proceedings of the IEEE International Conference on Neural Networks. Piscataway, NJ, 1942-1948.

[5] Yang, X. S. 2008. Nature-inspired Metaheuristic Algorithm. Luniver Press.

[6] Chen, T., Zhang, B., Hao, X., and Dai, Y. 2006. Task Scheduling in Grid based on Particle Swarm Optimization. In Proceedings of the Fifth International Symposium on Parallel and Distributed Computing (ISPDC'06). 238-245.

Scalable Algorithm for High Utility Itemset Mining

Chithra Ramaraju
Research Scholar
Department of Computer applications,
National Institute of Technology, Tiruchirappalli-620015
chithra_viswanathan@yahoo.co.in

Nickolas Savarimuthu
Associate Professor
Department of Computer Applications,
National Institute of Technology, Tiruchirappalli-620015
nickolas@nitt.edu

ABSTRACT
High utility itemset mining has emerged as one of the active research area in data mining. Numbers of algorithm were proposed for mining high utility itemsets, using complex data structures for reducing number of candidate itemsets. In this paper, an efficient algorithm is proposed using vertical data representation for maintaining candidate itemsets, and transaction identifiers (TIDs) in which corresponding itemset appears. The main advantage of the vertical format is to support for fast calculation of utility value via intersection operations on TIDs and scalability with increased number of transactions and dimensions of the database. The experimental results of the proposed algorithm on different datasets show that the proposed algorithm outperforms Two-Phase algorithm especially for sparse datasets.

Keywords
High utility itemset mining, vertical data representation, sparse dataset.

1. INTRODUCTION
One of the important descriptive data mining techniques is traditional frequent pattern mining, which is a fundamental task for several knowledge discovery processes, such as association rule mining, sequential patterns, classification and clustering, time-series mining, graph mining, web mining and so on. The problem of mining frequent itemsets is to find complete set of itemsets in a transactional database (TDB) with respect to the given minimum support threshold. The initial solution to frequent pattern mining, the Apriori algorithm was proposed to generate all frequent itemsets and association rules for retail organization [1][2]. This algorithm adopts efficient candidate generation-and-test methodology by employing downward closure property for pruning infrequent itemsets at each level. At the same time, numbers of algorithms were proposed using vertical data representation for extracting frequent itemsets. Eclat [3] was the first algorithm to find frequent patterns by depth-first search and performed well. In vertical data format, each candidate itemset is associated with list of transactions (TIDs) in which it occurs, and lattice concept is used for decomposing equivalence candidate itemsets, in order to process them independently [4]. The VIPER [5] used compressed vertical bit vectors to represent transaction list of each itemset instead of TIDs. dEclat [6] stored the difference of TIDs called diffset between a candidate k itemset and its sub-sets, instead of intersecting TIDs sets. This algorithm computes support by subtracting the cardinality of diffset from the support of its subsets. MAFIA algorithm [7] used depth-first search to traverse itemset lattice with effective pruning mechanisms that significantly improve mining performance.

The frequent itemset mining model has two limitations. First it treats all items in the database equally with same importance /weight/preference. Secondly, the model considers only presence or absence of items in the transaction. Therefore, traditional frequent pattern mining cannot satisfy the requirement of finding the most valuable itemsets.

This gives motivation to develop a utility mining model to discover itemsets, which contribute to business organization with high profit. Utility mining framework was [8] proposed to identify itemsets in a transaction database whose utility value is higher than or equal to minimum threshold value specified by user. The same author [9] proposed two utility mining algorithms namely UMining and Umining_H. Two-Phase algorithm proposed in [10] is the best known algorithm which uses efficient candidate generation procedure. This algorithm defines weighted transaction utilization measure to prune down number of candidate itemsets in order to reduce search space. The Novel tree-based candidate pruning technique HUC-Prune (high utility candidates prune) is proposed in [11] to efficiently mine high utility patterns without level-wise candidate generation. WIT tree data structure proposed in [12] where each vertex includes itemset, Tidset and twu value. In phase I, All high transaction weighted utilization itemsets are discovered and in phase II, utility value of all high transaction weighted utilization itemsets are calculated. This approach requires complex tree structure and huge memory for maintaining TIDs of all candidate itemsets during mining process. In [13] utility mining with the aid of a HUP tree structure is proposed and HUP growth approach is used for extracting high utility patterns.

The challenge of high utility itemset mining is to restrict number of candidate itemset generation and simplifying the computation for calculating utility value. In order to tackle these challenges, new algorithm is presented in this paper, which uses simple data structure for discovering high utility itemsets. For generating itemset of k+1 length, itemset of k length candidate list is sufficient. The main contribution of this work is

1. Vertical database representation is used, where each candidate itemset is associated with a list of transactions (TIDs) in which it occurs. All high utility itemsets are extracted via simple TID-list intersections and computing utility value for high transaction weighted utility itemset leads to simpler implementation.
2. Pruning strategy is used to restrict the number of candidate sets further.

The rest of this paper is organized as follows: In Section 2, high utility pattern mining problem is described. In Section 3, lattice-based candidate itemset generation is discussed. Section 4 describes proposed algorithm. In Section 5, experimental results are discussed.

2. HIGH UTILITY PATTERN MINING
The basic terms and formal definition of high utility itemset based on [8] are given below. Let $I = \{i_1, i_2, i_3 \ldots i_m\}$ be a set of items and $TDB = \{T_1, T_2, T_3 \ldots Tn\}$ be a set of transactions, where each transaction has a unique identifier (TID) and contains set of

purchased items in I. The purchased quantity of an item ip in a transaction Tq, is denoted by l(ip, Tq), is defined as numerical value stored in the transactional database shown in Table I(a). The profit per item ip is denoted by e(ip) in profit table.

Definition 1: The utility of an item ip in transaction Tq, is the quantity measure de-noted by U (ip, Tq), where

$$U(i_p, T_q) = l(i_p, T_q) \times e(i_p) \quad (1)$$

Definition 2: The **utility value** of an itemset X in the database U(X), is given as

$$U(X) = \sum_{i_p \in X} \sum_{T_q \in D} U(i_p, T_q) \quad (2)$$

In Two-Phase algorithm [10] transaction utility (tu), transaction weighted utility (twu), and high transaction weighted utility itemsets (htwu) are defined.

Table 1(a). Transaction Table

Table 1(b). Profit Table

Item	a	b	c	d	e
Profit	5	6	1	9	2

Definition 3: The **transaction utility** of transaction Tq, denoted as tu (Tq), is the sum of the total profit of all items in Tq and it is defined

TID	a	b	c	d	e	(tu)	Revised tu
1	10	1	0	5	6	113	113
2	4	5	0	7	1	115	115
3	0	1	13	0	2	23	10
4	6	5	0	0	10	80	80
5	7	1	0	4	3	83	83
6	0	4	10	0	0	34	24
7	7	3	0	0	0	53	53
8	0	3	0	2	7	50	50
9	4	0	0	1	0	29	29
10	0	0	0	5	12	69	69
total	38	23	23	24	41	649	

$$tu(T_q) = \sum_{i_p \in T_q} U(i_p, T_q) \quad (3)$$

4: Transaction weighted utilization of an itemset X, denoted by twu(X), is the sum of transaction utilities of all transactions containing X.

$$twu(X) = \sum_{X \subseteq T_q \wedge T_q \in D} tu(T_q) \quad (4)$$

Definition 5: The **minimum utility threshold** is the user preferred percentile of total transaction utility value of the given database.

$$min\_util = \partial \times \sum_{T_q \in D} tu(T_q) \quad (5)$$

where ∂ is the user preferred percentage.

Definition 6: X is a **high transaction-weighted utility** itemset, if twu(X) >= min_util, otherwise it is a low transaction weighted utility itemset.

Definition 7: An itemset X is a **high utility itemset** if u(X) >= min_util. Finding high utility itemsets means determining itemset X having criteria u(X) >= min_util.

The problem of mining high utility itemsets, given a user-specified minimum utility threshold ∂ and TDB, is to find the complete set of high utility itemsets in TDB with respect to a given minimum utility threshold. X is called a high utility itemset or high utility pattern if $U(X) \geq \partial$.

3. INTERSECTION OF TRANSACTIONS BASED ON PREFIX SHARING SUBSETS

Let I= $\{i_1, i_2, i_3 \ldots i_n\}$ be a set of n items from given transactional database. All possible combination of candidate itemsets are represented as power set lattice P(I) = 2^n-1 as shown in Figure. 1. High transaction weighted utility itemsets and high utility itemsets are shown in dark color.

Each item X of power set ($X \in I$) is associated with TIDs list denoted T(X), not shown for simplicity. For example, item A purchased in Transactions $\{T_1, T_2, T_4, T_5, T_7, T_9\}$ and item B purchased in $\{T_1, T_2, T_3, T_4, T_5, T_6, T_7, T_8\}$. New itemset of length k+1 can be generated by combining two items of the lattice, which must share k-1 length prefix. The transaction set of new itemsets can be obtained by intersecting TIDs list of its two subsets. The transaction lists of AB = T(A)∩T(B) = $\{T_1,T_2,T_4,T_5,T_7\}$. The twu value of AB is calculated from transaction set, and if it satisfies minimum utility threshold, the utility value of itemset is calculated by accumulating all the items value in each intersecting transactions.

In terms of data requirement, it is sufficient to store intermediate TIDs -list of at most two consecutive levels and candidate 1 item list; hence less memory is needed for mining high utility itemsets. Once candidate list for next level is generated, the candidate itemsets of current level can be deleted from the memory. In each level, candidate itemsets are generated partially in main–memory, when limited memory is available.

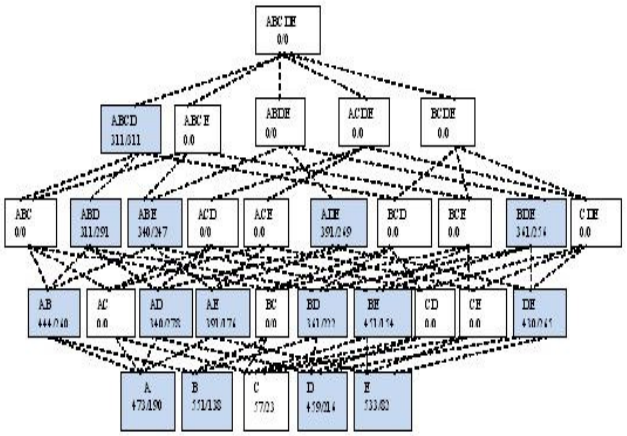


Figure 1. Power set lattice

Vertical Depth First Mine (VD-Mine): Some frequent pattern mining algorithms and WIT tree (Le et al., 2009) used depth wise strategy to traverse the lattice. The advantages of this traversing method is to discover a particular item and all its which contribute to overall profit and pruning strategy can be incorporated to reduce number of candidate itemsets. This strategy starts to process itemsets A, AB, AC, AD, ABC, ABD, ACD, and ABCD in the first level and B, BC, BD, and BCD in the second level followed by C, CD and D. While processing in the second level, item A's contribution can be reduced from transaction set of B, C and D. Similarly in the third level, item B's contribution can be reduced from transaction set of C and D. This is repeated for all levels. This pruning strategy reduces number of candidate sets in the subsequent level.

Pruning Strategy: Once particular item A and its combinations are discovered, the item A contributions can be reduced from transaction set of remaining items.

4. PROPOSED ALGORITHM DESIGN AND IMPLEMENTATION

All existing high utility pattern mining algorithms have utilized traditional horizontal transactional database, consists of list of transactions, where each transaction has an identifier followed by list of items and corresponding purchased quantities as shown in Table 1. The profit per item is given in a separate table. The proposed algorithm uses vertical database representation shown in Table 2. and contains itemset (htwu) followed by transaction TIDs. This representation facilitates fast discovery of itemset utility value by simple intersection of transaction TIDs and adding required utility values in each intersecting transactions. The depth first traversal (VD-Mine) algorithm is described in the following steps.

1. First step involves the computation of profit value of each item $U(i_p)$, by multiplying purchased quantity with the corresponding profit value in each transaction while reading database into main memory. The transaction utility value for each transaction tu (Tq) is calculated by accumulating utility value of each item in Tq. From the total database utility, minimum utility threshold is set for further processing.
2. Transaction weighted utility value of all item twu ($_{ip}$) is calculated and utility value is calculated only for high twu items. According to intermediate downward closure property, the low twu item and its superset cannot be high utility itemset, hence pruned from the candidate list. In this example, item 'C' is a low twu item and hence pruned from the table by reducing its utility value from all the transaction which contains "C" as shown in Table 1.
3. For each item in level 1, this step involves the generation of candidate k length candidate itemsets by combing item with other remaining items recursively in depth wise manner until there are no more itemsets are to be generated. The transaction set of k itemset is derived by intersecting TIDs of its subsets and twu value is computed. The utility value of k itemset is calculated by accumulating utility value of all items in the intersecting transactions as shown in Table 2. This process is repeated for all items in the first level.

Table 1. Candidate 1 itemset in vertical format

Items	Transaction	TWU	Util value
A	T1,T2 ,T4,T5,T7,T9	473	190
B	T1,T2,T3,T4,T5,T6,T7,T8	528	138
D	T1,T2 ,T5,T8,T9,T10	459	216
E	T1,T2,T3,T4,T5,T8,T10	520	82

Table 2. A's related candidate itemset in vertical format

Items	Transaction	TWU	Util value
AB	T1,T2 ,T4,T5,T7	444	260
AD	T1,T2,T5,T9	340	278
AE	T1,T2,T4,T5	391	176
ABD	T1,T2,T5	311	291
ABE	T1,T2,T4,T5	340	247
ADE	T1,T2,T5	391	269
ABDE	T1,T2,T5	311	311

4. After generating k+1 length itemsets, k length itemset candidate lists can be removed from the main memory. For example, once ABD, ABE, ADE itemsets are generated, the itemsets AB,AD, AE are not required. Hence the algorithm scalable, with increased number of transactions and dimensions.

The pseudo code of the proposed algorithm is given as follows.

Algorithm. VD-HU-Mine

Input: (1) Transactional database (*DB)*

(2) Profit Table

(3) Min_util*:* minimum utility threshold value

Output: All High utility itemsets

depth-first-search (S)

{

for all single items $A_i \in S$

for all items $A_j \in S$ and j > i do

$I = A_i \cup A_j$ // itemset generation

```
    T(I) = T(Ai) ∩ T(Aj)// intersecting transaction subsets
        calculate twu(I)
     if (twu(I) > min_util)  Add it to Ci
        {
        calculate utility(I)  in each intersecting transactions
        if(u(I)  > min_util) Add it to Ui
        }
   end

    for all Ci ≠ ∅  do depth-first-search(Ci)
end
}
```

5. EXPERIMENTAL RESULTS

The performance of the proposed algorithm (VD-Mine) is compared with Two-Phase algorithm, which is downloaded from (CUCIS, North Western University website). Programs for proposed algorithm were written in C# and experimented on Intel dual core 3.00 GHz processor with 1.96 GB RAM, running the Windows XP operating system. Dataset Retail is available from FIMI repository and this dataset does not provide profit values and purchase quantity of each item in the transaction. For each item, purchased quantity is assigned with random numbers with range 1 to 20 and profit value between ranges 0.1 to 10. Chain store is available in a bench-mark suite (NU-Mine Bench 2.0). The characteristics of both datasets taken for experimental analysis are given below in Table 4. Experimental results of Figure. 2 (a) and (b) show the execution time obtained by Two-Phase and proposed algorithm (VD-Mine).

Table 4.Characteristics of Datasets

Datasets	Nature	No of Trans	No of Items	Average Length
Retail	Sparse	88162	16470	10.3
Chain-store	Sparse	1, 112, 949	46086	7.2

6. CONCLUSION

In this paper, new algorithm for fast vertical mining of high utility itemsets is presented which uses vertical data representation and simple data structure for discovering all high utility itemsets. Each candidate itemset is associated with a list of transactions (TIDs). All high utility itemsets are extracted via simple TIDs-list intersections and computing utility value lead to simpler implementation. Two search strategies were discussed for traversing the power set lattice in breadth wise and depth wise manner. The proposed algorithm uses depth first strategy, which takes only a few database scans.

The performance of the proposed work is evaluated and results are compared with existing benchmark algorithm for three different datasets. The experimental results show that the proposed algorithm outperforms Two-Phase algorithm in terms of execution time and memory requirement especially for sparse datasets.

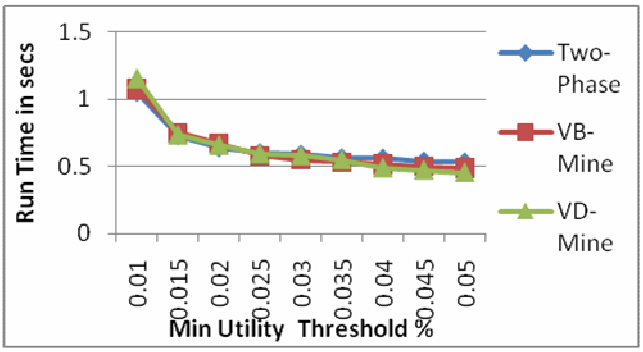


Figure 2(a). Runtime comparison on the Retail dataset

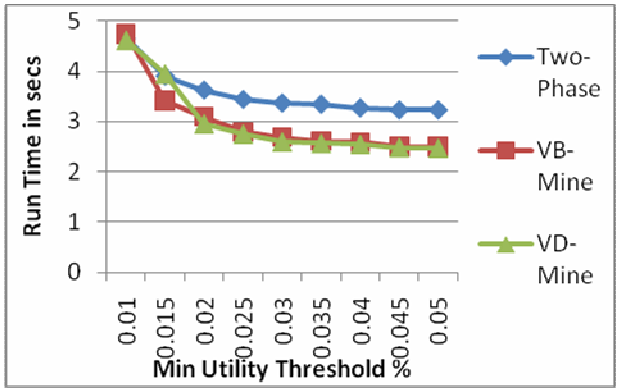


Figure 2(b). Runtime comparison on Chain-store dataset

7. REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A. 1993. Mining Association Rules between Sets of Items in Large Databases. *ACM SIGMOD International Conference on Management of Data.* May, 1993, pp. 207-216.

[2] Agrawal, R., Srikant, R. 1993. Fast Algorithms for Mining Association Rules. *Proceedings 20th International Conference on Very Large Data Base.* Sept, 1994, pp. 487-499.

[3] Zaki, M.J., Parthasarathy, S., Ogihara, M., and Li, W. 1997. New Algorithms for Fast Discovery of Association Rules. *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, AAAI Press.* 1997, pp. 283-286.

[4] Zaki, M.J. 2000. Scalable Algorithms for Association Mining. *IEEE Transactions on Knowledge And Data Engineering.* VOL. 12, NO. 3, MAY/JUNE 2000.

[5] Shenoy, P. Haritsa, J. R. , Sudarshan, S., Bhalotia, G., Bawa, M. and Shah, D. 2000 Turbo- Charging Vertical Mining of Large Databases. *Proceedings of ACM SIGMO.,* May 2000 pp. 22-23.

[6] Zaki, M.J., and Gouda, K. 2003. Fast Vertical Mining Using Diffsets, *Proceedings of the Nineth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Washington, D.C., ACM Press, New York, pp. 326-335, 2003.

[7] Burdick, M. Calimlim, and Gehrke, J. 2001. MAFIA: a Maximal Frequent Item Set Algorithm for Transactional Databases. *Proceedings of International Conference on*

Data Engineering. Heidelberg, Germany, April 2001,pp. 443-452.

[8] Yao, H., Hamilton, H.J., and Butz, C. J. 2004. A Foundational Approach to Mining Itemset Utilities from Databases, *Proc of the fourth SIAM International Conference on Data Mining*. 2004, pp. 482-486.

[9] Yao, H., and Hamilton, H.J. 2006. Mining Itemset Utilities from Transactional Databases. *Data & Knowledge Engineering*. 59, 2006, pp. 603-626.

[10] Liu, Y., Liao W.K. and Choudhary A. 2005. A Two Phase Algorithm for Fast Discovery of High Utility Itemset. Cheng, D. and Liu, H. (eds). *PAKDD, LNCS (LNAI).* Vol 3518 (2005), pp 689- 695, Springer, Heidelberg.

[11] Ahmed, C.F., Tanbeer, S.K., Jeong, B.-S., and Lee, Y. An Efficient Candidate Pruning Technique For High Utility Pattern Mining. *Lecture Notes in Artificial Intelligence.* Springer, Vol 5476 , (2009) ,pp. 689- 695 ,Berlin, Germany.

[12] Le, B., Nguyun, H., Cao,T., Vo, B. 2009. A Novel Algorithm for Mining High Utility Itemsets. *In Proceedings of First Asian Conference on Intelligent Information and Database Systems.*

[13] Lin C.-W., Hong,T.-P., Lu W.-H. 2011. An Effective Tree Structure for Mining High Utility Itemsets. *Expert systems with its Applications.* 38(2011) 7419-7424.

[14] Frequent Itemset Mining Dataset Repository, http://fimi.cs.helsinki.fi/data/, accessed Jan. 2008.

[15] Pisharath, J., Liu, Y., Parhi, J., Liao, W.K., Choudhary, A., and Memik G. NU-MineBench Version 2.0 Source Code and Datasets, http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html

Comparison of Neural Network Models for Forecasting Daily Stock Price

Iman Sanjaya
Ministry of Communication and Information Technology
Jl.Merdeka Barat No.9 Jakarta,
Indonesia 10110
Telp. +62 21 34833640

iman.sanjaya@kominfo.go.id

ABSTRACT
This paper aims to compare the forecasting accuracy of Artificial Neural Network (ANN). Three different ANN models, namely Multilayer Perceptron (MLP), Radial Basis Function (RBF) neural networks, and Group Method of Data Handling (GMDH) Polynomial neural networks are used for the purpose of forecasting. Various combinations of data preprocessing techniques, network topologies, functions, and training algorithms are presented in order to obtain the best model. The empirical result shows that the GMDH Polynomial neural networks give the best forecasting accuracy.

Keywords
Forecasting, time series, neural networks, MLP, RBF, Polynomial.

1. INTRODUCTION
The ability to accurately predict the future is fundamental to many decision processes. Forecasting is an important subject that has been used in various sectors, especially in economic and financial disciplines. Many quantitative modeling tools have been invented in order to get a powerful forecasting method. One of the promising alternative tools is Artificial Neural Network (ANN) which has showed its superiority compared to the conventional linear models. Linear methods have dominated forecasting subject for many decades. Linear methods are easy to develop and implement; they are also relatively simple to understand and interpret. However, linear models have a serious limitation; they are not able to capture any nonlinear relationships in the data [1].

In this paper, three neural network models, - Multilayer Perceptron (MLP), Radial Basis Function (RBF) neural networks, and GMDH Polynomial neural networks - will be employed to time series analysis. Each model will be compared for demonstration of merits in terms of accuracy.

This research is conducted by studying corresponding literatures to understand the concept and theory of neural network. Real-world data will be applied to implement the neural networks forecasting, and DTREG software packages will be used to implement the proposed algorithms and GUI for time series analysis.

This paper contributes to provide comparison among three different neural network models together by using one time series data.

2. CONCEPT AND THEORY
2.1 Definition
Haykin [2] offered the following definition of a neural network: "A neural network is a massively parallel distributed processor made up simple processing units, which has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

1. Knowledge is acquired by the network from its environment through a learning process
2. Interneuron connection strengths, known as synaptic weights, are used to store the acquired knowledge."

2.2 Multilayer Perceptron neural network model
The following figure illustrates a perceptron network with three layers [3]:

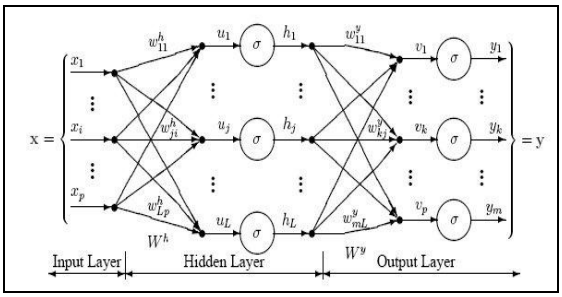


Figure 1. MLP architecture [3]

This network has an input layer (left) with three neurons, one hidden layer (middle) with three neurons and an output layer (right) with three neurons. There is one neuron in the input layer for each predictor variable $(x_1,...,x_p)$.

A vector of predictor variable values $(x_1,...,x_p)$ is presented to the input layer. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the bias, that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ji}), and the resulting weighted values are added together producing a combined value u_j. The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value h_j. The outputs from the hidden layer are distributed to the output layer.

Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value v_j. The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value y_k. The *y* values are the outputs of the network. If a regression analysis is being performed with a continuous target variable, there is a single neuron in the output layer generating a single *y* value.

2.3. Radial Basis Function (RBF) neural networks

RBF networks have three layers [3], as can be seen in Figure 2:
Input layer – There is one neuron in the input layer for each predictor variable. The input neurons (or processing before the input layer) standardize the range of the values by subtracting the median and dividing by the interquartile range. The input neurons then feed the values to each of the neurons in the hidden layer.

Hidden layer – This layer has a variable number of neurons (the optimal number is determined by the training process). Each neuron consists of a radial basis function centered on a point with as many dimensions as there are predictor variables. The spread (radius) of the RBF function may be different for each dimension. The centers and spreads are determined by the training process. When presented with the ***x*** vector of input values from the input layer, a hidden neuron computes the Euclidean distance of the test case from the neuron‘s center point and then applies the RBF kernel function to this distance using the spread values. The resulting value is passed to the summation layer.

Summation layer – The value coming out of a neuron in the hidden layer is multiplied by a weight associated with the neuron and passed to the summation which adds up the weighted values and presents this sum as the output of the network.

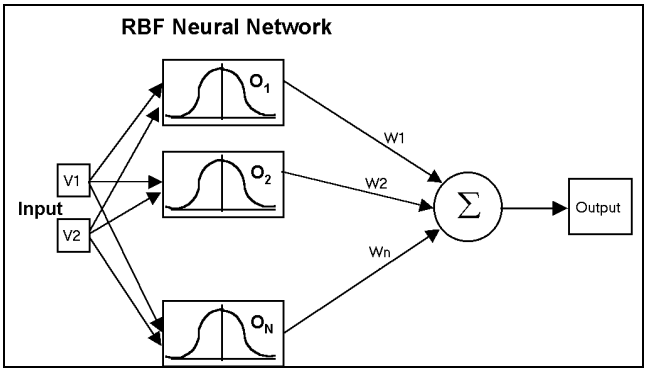


Figure 2. RBF neural network [3]

2.4. Polynomial Neural Networks (PNNs)

2.4.1. Group Method of Data Handling (GMDH) Polynomial neural networks

Group Method of Data Handling (GMDH) polynomial neural networks are “self organizing” networks. The network begins with only input neurons. During the training process, neurons are selected from a pool of candidates and added to the hidden layers.

2.4.2. Structure of a GMDH network

GMDH networks are *self organizing*. This means that the connections between neurons in the network are not fixed but rather are selected during training to optimize the network. The number of layers in the network is also selected automatically to produce maximum accuracy without overfitting.

The following figure from Kordik, Naplava, Snorek in [3] illustrates the structure of a basic GMDH network using polynomial functions of two variables:

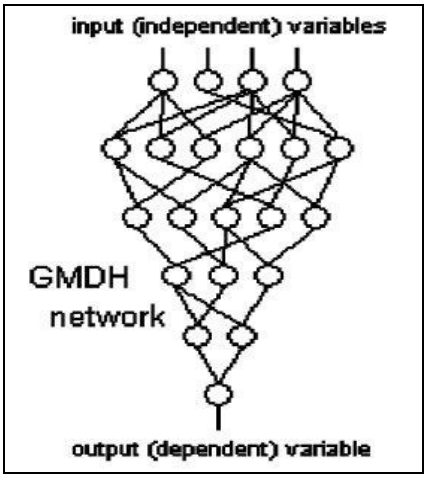


Figure 3. Basic GMDH network using polynomial function of two variables [3]

The first layer (top) presents one input for each predictor variable. Each neuron in the second layer draws its inputs from two of the input variables. The neurons in the third layer draw their inputs from two of the neurons in the previous layer; this progresses through each layer. The final layer (bottom) draws its two inputs from the previous layer and produces a single value which is the output of the network.

3. DESIGNING NEURAL NETWORKS FOR FORECASTING DAILY STOCK PRICES

3.1 Experimental setting

In this section, experiments using real financial data are conducted to show the neural networks capabilities for financial data prediction. Three different types of neural networks - Multilayer Perceptron Neural Networks, Radial Basis Function Neural Networks, and Polynomial Neural Networks - are applied to obtain the best model. Various combinations of data preprocessing, network architectures/topologies, activation functions, and training algorithms are compared for each neural network model.

This paper covers the time period of January 12, 2007 through July 2, 2010. Thus, the data set contained 835 data points in the

time series. We used data from PT.Indosat, Tbk which is the second largest mobile communication company in Indonesia. The data consists of daily closing stock prices obtained from Yahoo Finance (http://finance.yahoo.com/). The start date chosen is January 12, 2007 since on that day Qatar Telecom started to take over 25% of Indosat's shares from STT AMH.

3.2 Input variable selection

The first step is selecting the input variables. Selecting input variables for the neural networks model is a critical factor for the performance of the neural networks. For time series forecasting problems, the inputs are typically past observations of the data series and the output is the future value. Since this paper uses univariate time series model, the selected input variables for this model are the lagged observation of the time series being forecasted. The ANN performs the following function mapping

$$y_{t+1} = f(y_t, y_{t-1}, ..., y_{t-p}) \quad (1)$$

where y_t is the observation at time t. Thus the ANN is equivalent to the nonlinear autoregressive model for time series forecasting problems [4].

3.3 Data preprocessing

In some literatures, in order to reduce the amount of abnormal data and to increase convergence speed of the network, before any data is used in the training, the data is often scaled/normalized. For example, the daily stock prices could be divided by the maximum historical price (e.g. 9,900 in this time series data). However, in this paper, data scaling/normalization is not necessary. According to Zhang [5], there is no consensus on whether data normalization should be used. For example, it is still unclear whether normalizing the inputs is necessary because the arc weights can undo the scaling. Shanker, et al, in Zhang [5], investigated the effectiveness of linear and statistical normalization methods for classification problems. They found that data normalization does not necessarily lead to better performance particularly with large networks and samples. Zhang [5] found no significant difference between using normalized and original data. Hence, in this paper, instead of normalizing data, a trend removal is used in data preprocessing (and may be followed by variance stabilization). Raw data is also used in the experiments since Dorffner [6] and Balkin [7] suggested that stationarity assumption is not needed and applying differencing in ANN forecasting is not necessary for neural networks.

3.4 Datasets partitioning

The datasets are divided into training sets and validation sets. Training sets start from the beginning of the data to June 11, 2010, and the rest of data are used for validation sets. It is based on practical guide from Kaastra [8] that the training set should be the largest set and is used by the neural network to learn patterns present in the data, while the validation set should consist of the most recent contiguous observations. The validation set is used to select the optimal network parameters for the model. Training and validation occur simultaneously, and the two sets of data are used for exploration of parameters value of the model.

3.5 Multilayer Perceptron Model for stock price prediction

3.5.1. Number of input

The criticality in selecting the input variables lies in selecting the number of lagged variables of the input variables. Balkin [9] recommended that the possible lags that should be chosen for daily data are lags 1 to 6. Therefore, the lags 1 to 6 are chosen as input variables in this paper, and we have

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}, y_{t-6}) \quad (2)$$

3.5.2. Number of hidden layers

DTREG supports up to two hidden layers, therefore one or at most two hidden layers are used in the experiments to find the best model, and this is in line with Kaastra [8] who recommended that all neural networks should start with preferably one or at most two hidden layers. In practice, neural networks with one and occasionally two hidden layers are widely used and have performed very well. Increasing the number of hidden layers increases computation time and may result in overfitting which leads to poor out-of-sample forecasting performance.

3.5.3. Number of hidden neurons

This research carries out an exhaustive search method among all topologies of the search range we specified, up to 2 hidden layer networks, to determine the best topology of the network.

3.5.4. Number of output

Deciding on the number of output is more straightforward since there are compelling reasons to always use only one output neuron. Neural networks with multiple outputs, especially widely spaced outputs, will produce inferior results compared to a network with a single output [4].

3.5.5. Activation/transfer function

While the majority of researchers use logistic (sigmoid) activation functions for hidden nodes, there is no consensus on which activation should be used for output nodes. A number of researchers simply use the logistic (sigmoid) activation functions for all hidden and output nodes [4]. In this paper, combinations of linear and logistic (sigmoid) activation functions are used both for hidden and output layers.

3.5.6. Evaluation criteria

Four most frequently used evaluation criteria, namely ME (Maximum Error), MSE (Mean Square Error), MAE (Maximum Absolute Error) and MAPE (Mean Absolute Percentage Error) are used.

3.5.7. Training algorithms

Two training methods are used, i.e., scaled conjugate gradient and traditional conjugate gradient.

3.5.8. Experimental result

The MLP model which produces the minimum forecast error in the validation set is selected. A total of 48 possible models are used in this experiment. We get model #27 as the best model for the MLP because it gives the lowest value of three different criteria for the validation set. Model #27 can be described as follows:

- Data preprocessing technique in the form of trend removal without variance stabilization is used for the data. Exponential trend is chosen as following equation (3):

$$price = 88670.965824 - 81455.196630 \; x \exp(0.000034 \; x \; row)$$

- One hidden layer
- Logistic (sigmoid) activation functions are employed in the hidden layer and the linear activation function is utilized in the output layer
- Traditional conjugate gradient as the training method
- The selected best model uses 6-8-1 as the network topology (6 input nodes, 8 hidden neurons, and one output node).

The time series chart is displayed in Figure 4. A black square indicates the actual value of the time series, a green square indicates the predicted value for points corresponding to training points, and an open blue circle indicates the predicted value for validation rows not used for training.

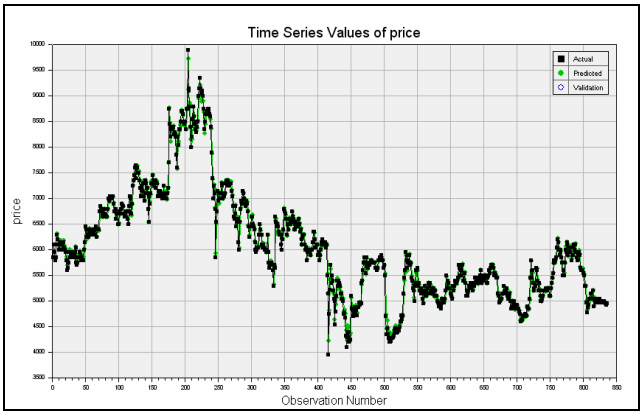


Figure 4. Time series values of price for MLP (6-8-1)

The actual versus predicted chart (residual chart) is shown in Figure 5. It displays a point for each data row. The X coordinate of a point is the actual target value. The Y coordinate of the point is the corresponding predicted target value. In a perfect model, the predicted values would equal the actual values, the X and Y coordinates for each point would be equal, and all points would be located on the diagonal line where X = Y. When the predicted value differs from the actual value, the points are offset from the diagonal line, and the vertical distance from the line to the point corresponds to the error (residual). The error is denoted by red vertical lines.

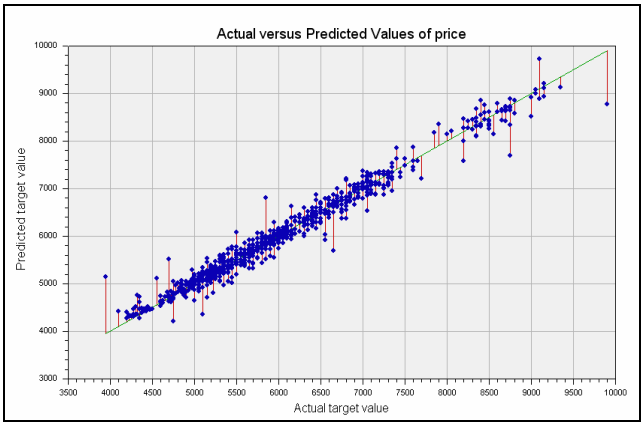


Figure 5. Actual versus predicted values of price for MLP (6-8-1)

3.6 Radial Basis Function (RBF) networks model for stock price prediction

3.6.1. Variable Input and Data Preprocessing

In order to have a realistic comparison, the RBF neural network is also modeled with six inputs. Three different data preprocessing processes are proposed in this paper, namely:

- RBF model #1: using raw data
- RBF model #2: using trend removal without variance stabilization
- RBF model #3: using trend removal with variance stabilization

3.6.2. Training RBF networks

DTREG uses a training algorithm developed by Sheng Chen, Xia Hong and Chris J. Harris (2005). This algorithm uses an evolutionary approach to determine the optimal center points and spreads for each neuron. It also determines when to stop adding neurons to the network by monitoring the estimated leave-one-out (LOO) error and terminating when the LOO error beings to increase due to overfitting [3].

The computation of the optimal weights between the neurons in the hidden layer and the summation layer is done using ridge regression. An iterative procedure developed by Mark Orr (1966) is used to compute the optimal regularization Lambda parameter that minimizes the generalized cross-validation (GCV) error [3].

3.6.3. Empirical result

RBF model #3 is chosen as the best model since it results in the minimum forecasting error (see Table 1). The number of units (the number of RBFs) for the single hidden layer is 26; the minimum spread parameter (the width of RBFs) is 0.01 and the maximum is 547.813. The minimum and maximum Lambda parameters are, respectively, 0.10605 and 9.5768.

Table 1. Validation result for all possible RBF models.

RBF Model	Validation Data			
	ME	MSE	MAE	MAPE
RBF Model #1	232.94177	18483.349	118.27438	2.380250
RBF Model #2	282.04793	42553.881	194.59121	3.903748
RBF Model #3	82.24053	969.867	21.29296	0.425042

3.7 GMDH Polynomial Neural Networks model for stock price prediction

In this section, three different classes of polynomials, namely linear, quadratic, and cubic are utilized and combined with three different data preprocessing techniques, which result in 9 proposed models. The functions used in the network can be seen in Table 2, while the preprocessing techniques are similar to the previous NN models.

3.7.1. Network parameters

Some network parameters must be specified before the training process.

a. Maximum network layers:
The maximum number of layers in the neural network that the model may contain is specified as 20. The model may actually be created with fewer layers if the building process discovers that adding layers would harm or not improve the accuracy of the model.

b. Maximum polynomial order:
The highest power of a variable that a polynomial may contain is specified as 16. If the GMDH network is built using quadratic polynomials, the order of the polynomials doubles on each layer.

Table 2. Functions to use in the GMDH network

No.	Functions	Description
1	Linear : 1 variable	$y = p_1 + p_2 x_1$
2	Quadratic : 1 variable	$y = p_1 + p_2 x_1 + p_3 x_1^2$
3	Cubic : 1 variable	$y = p_1 + p_2 x_1 + p_3 x_1^2 + p_4 x_1^3$

c. Convergence tolerance:
The training algorithm will add layers to the network until the specified convergence tolerance is reached or it is stopped for another reason, such as reaching the maximum allowed number of layers or detecting that adding a layer will not improve the model. The convergence tolerance value specifies the proportion of residual unexplained variance that is left, which is the remaining R^2 variance. For example, if a tolerance factor of 0.001 is specified, the algorithm iterates until the residual, unexplained R^2 reaches 0.001, which means the explained R^2 reaches 0.999 (99.9%). In this paper a tolerance factor of 0.0001 is used.

d. Number of neurons per layer:
The number of neurons in each layer of the network is specified as the same number of neurons in the input layer.

e. Network layer connections:
This parameter controls how neurons in the network are connected together. In this paper, only the type of connection to the previous layer is chosen. This tells DTREG that the inputs to one layer may come only from outputs generated by the next lower layer.

3.7.2. Empirical result

The training result is shown in Table 3. Based on minimum error performance in validation sets, model #4 is chosen as the best model for GMDH Polynomial NN. Model #4 can be described as follows:

- Trend removal (without variance stabilization) is used for the data. Exponential trend is exactly the same as equation (3).
- Linear function is used in the network
- The resulted GMDH model is as follows:

*N(2) = -0.701019+0.041907*price_Lag_6+0.928988*price_Lag_1*
*N(1) = -2.491868e-013+1*N(2)*
*N(6) = -1.209706+0.966447*price_Lag_1*
*N(5) = -2.092394e-013+1*N(2)-8.016434e-015*N(6)*
*N(8) = -1.199916+0.003746*price_Lag_2+0.962842*price_Lag_1*
*N(7) = 2.511918e-013+1*N(8)*
*price = -2.085145+7.07601e+013*N(1)-7.07601e+013*N(5)-0.627353*N(7)*

Output from neuron i is shown as $N(i)$. Each line above is referred as Partial Description. The final line shows the output of the network. It can be seen that the network consists of three layers.

Table 3. Validation result for 9 possible models of polynomial neural networks

GMDH Model #	Validation Data			
	ME	MSE	MAE	MAPE
1*	220.84	15410.200	105.960	2.133
2*	173.93	8651.220	80.317	1.616
3*	201.38	13037.000	98.701	1.987
4**	51.01	592.462	18.911	0.378
5**	72.98	1400.110	29.861	0.597
6**	85.75	2936.230	47.569	0.953
7***	56.44	690.386	20.044	0.401
8***	73.32	1457.440	30.833	0.616
9***	62.99	756.228	20.071	0.401

* Using raw data
** using trend removal
*** using trend removal and variance stabilization

4. DISCUSSION

Each neural network model has some specific factors that may influence the model accuracy. Multilayer perceptron neural networks rely on some factors, such as input variable(s), data preprocessing, datasets partitioning, network topology (number of input, number of hidden layer, number of hidden neuron, and number of output), the choice of activation/transfer function, and training algorithm. In this paper, the best selected model uses 6-8-1 topology (6 input nodes, 8 hidden neurons, and 1 output node), where logistic (sigmoid) functions are employed in the hidden layer and the linear function is utilized in the output layer. The network is trained using traditional conjugate gradient algorithm after implementing an exponential trend removal (without variance stabilization) in data preprocessing.

Radial basis function neural networks rely on some network parameters such as radius and Lambda. The best RBF model in this experiment comes with 26 neurons in a single hidden layer. The spread parameters are laid between 0.01 and 547.813, while the Lambda parameters are laid between 0.10605 and 9.5768. Exponential trend removal with variance stabilization is chosen as the data preprocessing technique.

GMDH polynomial neural networks rely on the determination of maximum number of layers, maximum polynomial order, function used in the network, number of neurons per layer, and network layer connection. In this paper, linear functions seems better than quadratic or cubic functions. A similar data preprocessing in multilayer perceptron model is used.

In this research, the greatest accuracy among three neural network models comes from GMDH Polynomial Neural Network. It may be caused by the self-organizing approach which is substantially

different from deductive methods used commonly for modeling. GMDH Polynomial Neural Network has inductive nature to find the best solution by sorting-out of possible variants. The interested reader may want to refer to the paper by Ivakhnenko [10] for the original concept of GMDH.

Table 4. Comparison of forecast accuracy among time series models

Time Series Model	Validation Data		
	MSE	MAE	MAPE
MLP	670.89	19.76	0.395
RBF	969.87	21.29	0.425
GMDH Polynomial	592.46	18.91	0.378

5. CONCLUSION

5.1 Conclusion

This paper compared three neural network models (multilayer perceptron, radial basis function neural networks, and GMDH polynomial neural networks) to forecast daily stock prices. Daily closing price data were used from 12, 2007 through July 2, 2010.

Results showed that the GMDH polynomial neural network forecasts were considerably more accurate than the other models. The mean squared error, mean absolute error, and mean absolute percentage error of the GMDH polynomial neural network models forecast were lower than those of the other models. However, we may need more sets of time-series data to ensure the drawn conclusion.

5.2 Recommendation

The research method suffers from a number of limitations. First, despite the satisfactory performance of the NN models in this paper, future research might improve the performance of the NN models used in this paper by using a hybrid approach, i.e., integrating certain NN model with stochastic model, fuzzy model, wavelet or genetic algorithm. Second, future research might use other learning algorithms beside scaled/traditional conjugate gradient, such as quick propagation, quasi Newton algorithm, and Levenberg-Marquardt to obtain the optimal result.

6. ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Hak-Keung Lam from King's College London, UK, who gave me the research topic and guided me with necessary research work.

7. REFERENCES

[1] Zhang, G. P. 2004. Neural Networks in Business Forecasting. Idea Group Publishing, USA

[2] Haykin, S. 1999. Neural Networks a Comprehensive Foundation, Second Edition, Prentice-Hall International, Inc. Upper Saddle River, N.J.

[3] Sherrod, P.H. 2010. DTREG – Predictive modeling software.

[4] Zhang, G., Patuwo, B. E. and Hu, M. Y. 1998. Forecasting with Artificial Neural Networks : the State of the Art. International Journal of Forecasting, vol.14, pp.35-62.

[5] Zhang, G. Hu, M. Y. 1998. Neural Network Forecasting of the British Pound/US Dollar Exchange Rate. Omega Int. J. Mgmt. Sci., vol 26, no. 4, pp.495-506.

[6] Dorffner, G. Neural Networks for Time Series Processing. Dept. of Medical Cybernetics and Artificial Intelligence, University of Vienna.

[7] Balkin, S.D. 1999. Stationarity Concerns when Forecasting using Neural Networks. INFORMS presentation, Philadelphia.

[8] Kaastra, I., Boyd, M. 1996. Designing a Neural Network for Forecasting Financial and Economic Time Series. Neurocomputing, vol.10, pp.215-236.

[9] Balkin, S. D., Ord, J. K. 2000. Automatic Neural Network Modeling for Univariate Time Series. International Journal of Forecasting, vol.16, pp.509-515.

[10] Ivakhnenko, A. G. 1970. Heuristic Self-Organization in Problems of Engineering Cybernetics. Automatica, Vol.6, pp.207-219.

The Approach for Table Extraction in Internet Based on Property and Instance

Detty Purnamasari
Information Technology, Gunadarma
Jl. Margonda Raya No. 100
Pondok Cina, Depok
+6221788811112 ex 206
detty@staff.gunadarma.ac.id

I Wayan Simri Wicaksana
Information Technology, Gunadarma
Jl. Margonda Raya No. 100
Pondok Cina, Depok
+6221788811112 ex 206
iwayan@staff.gunadarma.ac.id

Lintang Yuniar Banowosari
Information Management, Gunadarma
Jl. Margonda Raya No. 100
Pondok Cina, Depok
+6221788811112 ex 206
lintang@staff.gunadarma.ac.id

ABSTRACT
The Internet provides a source of data with a variety of forms and formats. One of the existing data is in tabular form. Tables can be retrieved and used for a further process. Taking a table from one source in the internet can be done by manually *copy-paste*, but it will be difficult if more than one table will be taken to merge of these tables. This paper discusses an algorithm for the extraction of tables in HTML format into a database form that have properties and records, and search properties of the extracted table. Then it is created a simple illustration to test the algorithm developed in this paper.

Keywords
HTML table, data capture, information query, table extraction

1. INTRODUCTION
Data can be obtained from various sources; one of them is the internet which provides data with a variety of forms and formats. Examples data forms which available on the internet is a table form, i.e. as the data available in the online shopping websites providing information about the item name, available item number, price, and so on. Retrieval of data in tabular form on the internet performed for processing the data to further process or to merge the extraction data with the data that already exists.

If you only take from one table and from one source then the *copy-paste* process is adequate. HTML table Extraction approach would be useful if taking more than one table from various sources on the Internet, the illustrations can be seen in Figure 1. The results of the data retrieval process of merging can be done to further process of interoperability.

Figure 1 shows two forms tables about ticket prices information with the property name is different but it has the same meaning (from = origin), which then the two values of tables are merged into one table only.

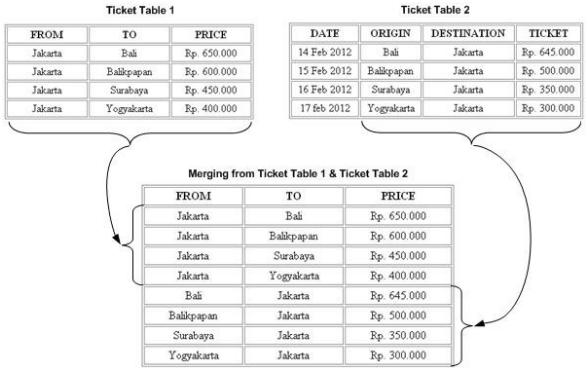


Figure 1. Ticket Table Merging Illustration

The important thing in table extraction is to capture value of cell in appropriate location and value. This problem will be represented in record model of database, furthermore in object oriented model will called instance.

This research develops an algorithm to perform the extraction of a simple table in HTML format into a database forms taking into account the property and records factor. These algorithms consider how many lines become a property and records. On the contents of these tables, there are also merge line/row and merge column.

Lerman et al [5] in their research developed how to extract a table in the form of lists and tables by using Unsupervised Learning Algorithms to obtain the structure of lists and tables. The algorithm is analyzed from several web pages to find template, until finally able to extract the data into rows and columns. Whereas it can be a web only has one page and on this one page required the extraction of the contents of a list or table.

Extraction table with a tree representation based on the Document Object Model (DOM) is done in several studies including by Lin et. al. [6], Gultom et. al. [2] and Hong et. Al [3] that use for the extraction of HTML. Krupl et. al. [4] not only uses the DOM tree but also use the XY cut algorithm to find a table on a web page.

Embley et.al [1] do the extraction of an HTML table to get the schema matching to allow for merging the results with other data extraction. The problems found by Embley et. al [3] in extracting HTML tables, some of which is the location of the table on a web page, merged attributes / values, and synonyms if it will merge the data extraction result.

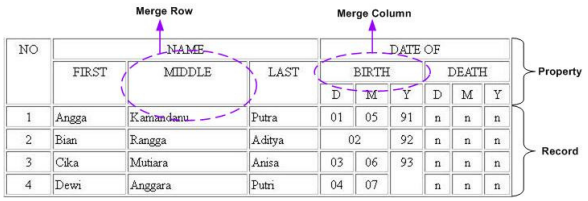
2. DESIGN OF TABLE EXTRACTION ALGORITHM

Tabular form with a simple HTML format that was developed in this study refine the approach to table 1 into a more complex table as in Table 2 that there is merging rows and columns with several rows of the property.

Table 1. Simple Table Form

NO	NAME	AREA CODE	PHONE
1	Ali	021	79090301
2	Bambang	022	31709010
3	Cepi	023	31901981

Table 2. Table Form with Merge Row, Merge Column, Property and Record



NO	NAME			DATE OF					
	FIRST	MIDDLE	LAST	BIRTH			DEATH		
				D	M	Y	D	M	Y
1	Angga	Kamandanu	Putra	01	05	91	n	n	n
2	Bian	Rangga	Aditya	02		92	n	n	n
3	Cika	Mutiara	Anisa	03	06	93	n	n	n
4	Dewi	Anggara	Putri	04	07		n	n	n

In the example for this algorithm, the row as the property is a first row up to third row, where its property there is merge rows and columns. In addition, the row of data record from that table there are also merge rows and columns.

Rows and columns merging process can be depicted using tree which show in figure 2.

Figure 3 shows the HTML tag for table 2 which shown first row until fifth row that we tested in this research.

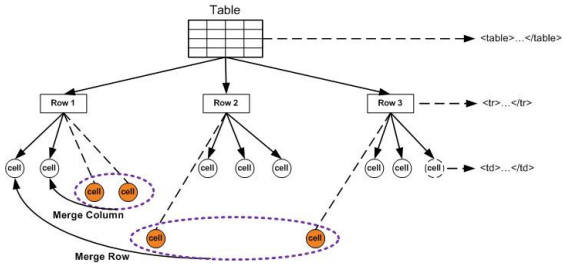


Figure 2. Tree for Merge Row and Column

```
<html>
<head>
  <title>tabel_ujicoba_ok</title>

  <meta http-equiv="content-type"
 content="text/html; charset=ISO-8859-1">
</head>
  <body>

<table cellpadding="2" cellspacing="2" border="1" width="50%">
   <tbody>
     <tr>
       <td valign="top" rowspan="3" colspan="1" align="center">NO<br>
       </td>
       <td valign="top" rowspan="1" colspan="3" align="center">NAME<br>
       </td>
       <td valign="top" rowspan="1" colspan="6" align="center">DATE OF<br>
       </td>
     </tr>
     <tr>
       <td valign="top" rowspan="2" colspan="1" align="center">FIRST<br>
       </td>
       <td valign="top" rowspan="2" colspan="1" align="center">MIDDLE<br>
       </td>
       <td valign="top" rowspan="2" colspan="1" align="center">LAST<br>
       </td>
       <td valign="top" rowspan="1" colspan="3" align="center">BIRTH<br>
       </td>
       <td valign="top" rowspan="1" colspan="3" align="center">DEATH<br>
       </td>
     </tr>
     <tr>
       <td valign="top" align="center">D<br>
       </td>
       <td valign="top" align="center">M<br>
       </td>
       <td valign="top" align="center">Y<br>
       </td>
       <td valign="top" align="center">D<br>
       </td>
       <td valign="top" align="center">M<br>
       </td>
       <td valign="top" align="center">Y<br>
       </td>
     </tr>
     <tr>
       <td valign="top" align="center">1<br>
      </td>
       <td valign="top">Angga<br>
      </td>
       <td valign="top">Kamandanu<br>
      </td>
       <td valign="top">Putra<br>
      </td>
       <td valign="top" align="center">01<br>
      </td>
       <td valign="top" align="center">05<br>
      </td>
       <td valign="top" align="center">91<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
     </tr>
     <tr>
       <td valign="top" align="center">2<br>
      </td>
       <td valign="top">Bian<br>
      </td>
       <td valign="top">Rangga<br>
      </td>
       <td valign="top">Aditya<br>
      </td>
       <td valign="top" align="center" rowspan="1" colspan="2">02<br>
      </td>
       <td valign="top" align="center">92<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
       <td valign="top" align="center">n<br>
      </td>
     </tr>
```

Figure 3. HTML Tag for Table 2

The tree shown in figure 2 described the first row to third row which is the property of table 2 with merging rows in row first to third and in column 1 and merging of columns 2, 3, and 4 in the first row. The tree consists of three levels, namely: tag <table> ...</ table> as the root of a single table, "Row to-1 up to-3: tag <tr> ...</ tr>" is the branch that make up the number of rows in the table, and "cell: <td> ...</ td>" as a leaf that forms the rows into columns.

If the merging process of the 1st, the 2nd, and 3rd row in the 1st column, then the 1st cell in 2nd row and 3rd row will be merged with the first cell in row 1.

However, the merging process of the 2nd, 3rd, and 4th column in 1st row , then the 3rd and 4th cell in the 1st row will be merged with the 2nd cell that residing in the same row.

The algorithms for table extraction have the following rules:

1. If the first tag found is <html> then it can be done the extraction.
2. All tags and strings before <table> tag is ignored because it is not part of the table.
3. Tag and the string that was found after <table> tag and before the </ table> tag is a composer of the table.
4. Tag and string after the </ table> tag is ignored because it is no longer part of the table.
5. Rows in the table begins with <tr> tag and ends with the tag </ tr>, and tag <td> ...</ td> make cell / column in the table.
6. Actual number of rows in the table can be determined by calculating the number of tags <tr> ...</ tr> tags that exist on <table> ... / table>. (*RsTotal* = number of rows).
7. The actual number of cell / column on the table can be determined by calculating the number of *colspan* values that exist in the first <tr> ...</ tr> tag. (*CsTotal* = number of columns).
8. Find the largest rowspan value of each tag <td> ...</ td> tags on <tr>...</ tr> to-s until was not found value of the *rowspan* > 1 to obtain the number of rows as property. (*Rowmax_pro* = boundary property line).
9. Taking the contents of the property, carried out starting from ending line of <tr> ...</ tr> tag which is the *rowmax_pro* down to 1 to obtain the position of cell / column if there is merging of column.
10. If the value *colspan* = 1, then the contents of cell / column directly taken and stored in the variable $\{TdVal\}_{(\text{row, cell position})}$.
11. If the value of *colspan* > 1, then searched the number and position of cell / column of each such member. Then the content of cell / column will run into merge / concat between the $\{\text{row}\}_s$ with the $\{\text{row}\}_{(s+1)}$ on $\{\text{cell position / column}\}_i$
12. Contents of records taken from the last row in the table / tag $\{\text{<tr> ...</ tr>}\}_{RsTotal}$ up to $\{\text{row}\}_{(rowmaxpro + 1)}$
13. There are three conditions, namely: i). If *colspan* = 1 and *rowspan* = 1, ii). If *colspan* = 1 and *rowspan* > 1, iii). If *colspan* > 1 and *rowspan* = 1.
14. If between tag <td> ...</ td> found "*rowspan*" variable means there are multiple rows that are merged into one, so the number of rows will be reduced as much as the *rowspan-1*.
15. If between tag <td> ...</ td> found "*colspan*" variable means there is some cell / columns that are merged into one, so the number of cells / columns will be reduced as much as the *colspan-1*.

Using those table extraction rules, then the algorithm developed for the extraction of the table with a few rows of property are as follows:

Algorithm 1. Count for Actual Number of Row and Actual Number of Column

```
Read HTML
s = 0
If read <tr> then s = s + 1
      RsTotal = s
Jum <td> = count tag <td>...</td>  in first tag <tr>...</tr>
CsTotal = 0
For i = 1 to jum<td>
      Read nilai cs (i)
      CsTotal = CsTotal + cs (i)
Next i ;
```

Notes :

- RsTotal : the total number of <tr> tag on tag <table>…</table>
- CsTotal : the number of colspan value
- Jum<td> : the number of <td> tag
- cs : colspan value
- i : <td> tag
- s : <tr> tag

In algorithm 1 is searched the actual number of rows (*RsTotal*) that obtained by reading the number of tags <tr> ...</ tr> which reside on <table> ...</ table> tag and the actual column (*CsTotal*) by counting the number of tags <td> ...</ td> that exist in first tags <tr> ...</ tr>. Then the actual number of columns and rows will be used in subsequent algorithms.

Algorithm 2 is used to find the value of the property rows boundary (*rowmax_pro*) by finding the largest *rowspan* value (*RsMax*) that exist in each tag $\{\text{<td> ...</ td>}\}_i$ on tag $\{\text{<tr> ...</ tr>}\}_s$. If it is not found any longer *rowspan* > 1 value, then the rows boundary of property was found.

Algorithm 2. Finding the Largest Rowspan Value, and Number of Row to be The Property Boundary

```
mBatas (0) = 1
while s = 1 do
    rsMax (0) = 1
    Count jum<td>
    For i = 1 to jum<td>
          If  rs (i) > 1 then
            If  rs (i) > = rsMax (i-1) then rsMax (i) = rs (i)
            else
            If rs (i) < rsMax (i-1) then rsMax (i) = rsMax (i-1)
     Next i ;
mBatas (s) = rsMax (i) + s – 1
if mBatas (s) < mBatas (s-1) then mBatas (s) = mBatas (s-1)
until mBatas (s) ;
rowmax_pro = mBatas (s) ;
```

Notes :

- RsMax : the highest value of rowspan
- mBatas : row value boundary as property
- rowmax_pro : row boundary which named as property
- i : <td> tag
- s : <tr> tag

Rows boundary of property (*rowmax_pro*) has been established, and then the contents of cell as a property name can be taken. The algorithm 3 begins by taking a cell value in the <td> ...</ td> tag which reside in the tags $\{<tr>...</ tr>\}_{(rowmax\_pro)}$ and the position of the column that will be used in storage variable $\{TdValue\}_{(s, PosisiCol\ (i))}$, where the contents of this cell will be merged forward if the tag <td> ...</ td> inside tag $\{<tr> ...</ tr>\}_{(rowmax\_pro\ -\ 1)}$ has a value *colspan* > 1.

Then by using the process of repetition (*rowmax_pro - 1*) down to (1) and see whether the two conditions *colspan* value of 1 or > 1 to determine the merge of the column, then the name of the property obtained.

Algorithm 3. Finding The Content of Property

```
s = rowmax_pro
For i = 1 to jum<td>
  TdVal (s, i) = value in tag {<td>…</td>}i
Next i;
For s = rowmax_pro – 1 down to 1
  PosisiCol (0) = 1
  jumAnggota (0) = 0
  c = 1
  For i = 1 to jum<td>
    If cs = 1 then
      PosisiCol (i) =
      (jumAnggota (i-1) + PosisiCol (i-1) – 1 ) + 1 ;
      jumAnggota (i) = cs (i)
      TdVal (s, PosisiCol(i)) = value in tag <td>…</td>
    If cs > 1 then
      PosisiCol (i) =
      (jumAnggota (i-1) + PosisiCol (i-1) – 1 ) + 1 ;
      jumAnggota (i) = cs (i)
      TdVal (s, i) = value in tag <td>…</td>
      For j = 0 to jumAnggota (i) - 1
        TdVal (s, PosisiCol(i) + j) =
        TdVal (s, i) concat TdVal (s + 1, c) ;
        c = c + 1
      Next j ;
  Next i;
Next s;
s = 1
For i = 1 to CsTotal
  Save TdVal (s, i) as property
Next i;
```

Notes :

- rowmax_pro : row boundary which named as property
- TdVal : variable to save tag <td>…</td> value
- PosisiCol : position of $\{column\}_i$ on $\{row\}_s$
- jumAnggota : number of column member if there is colspan value

Algorithm 4. Finding The Content of Record

```
s = RsTotal
For i = 1 to jum<td>
    TdVal (s, i) = value in tag {<td>…</td>}i
Next i;
For s = RsTotal - 1 down to rowmax_pro + 1
    PosisiCol (0) = 1
    jumAnggota (0) = 0
    For i = 1 to jum<td>
      If cs = 1 and rs = 1 then
        PosisiCol (i) =
        (jumAnggota (i-1) + PosisiCol (i-1) – 1 ) + 1;
        jumAnggota (i) = cs (i)
        TdVal(s,PosisiCol(i)) = value in tag <td>…</td>
      If cs = 1 and rs > 1 then
        PosisiCol (i) =
        (jumAnggota (i-1) + PosisiCol (i-1) – 1 ) + 1;
        jumAnggota (i) = cs (i)
        TdVal (s, i) = value in tag <td>…</td>
        While x = 1 do
          For j = rs + 1 down to 1
            TdVal (s + x, PosisiCol(i) + j) =
            TdVal (s + x, PosisiCol (i) + j – 1) ;
          Next j ;
          TdVal(s + 1, PosisiCol (i)) =TdVal (s, PosisiCol (i))
        Until x = rs - 1 ;
      If cs > 1 and rs = 1 then
        PosisiCol (i) =
        (jumAnggota (i-1) + PosisiCol (i-1) – 1 ) + 1 ;
        jumAnggota (i) = cs (i)
        TdVal (s, i) = value in tag <td>…</td>
        For j = 0 to jumAnggota (i) - 1
            TdVal (s, PosisiCol(i) + j) = TdVal (s, i)
        Next j ;
    Next i;
Next s;
For s = rowmax_pro + 1 to RsTotal
    For i = 1 to CsTotal
      Save TdVal (s, i) as property
    Next i ;
Next s ;
```

Notes :

- rowmax_pro : row boundary which named as property
- TdVal : variable to save tag <td>…</td> value
- PosisiCol : position of $\{column\}_i$ on $\{row\}_s$
- jumAnggota : number of column member if there is colspan value

Algorithm 4 is used to retrieve the contents of the tag <td> ...</ td> is inside tag <tr> ...</ tr> started in the last row of the table (*RsTotal*). Then repeat to retrieve the contents of records starting from (*RsTotal* - 1) down to (*rowmax_pro* + 1). In algorithm 4 is divided into three conditions, namely: i). If *colspan* = 1 and *rowspan* = 1; ii). If *colspan* = 1 and *rowspan* > 1, and iii). If *colspan* > 1 and *rowspan* = 1. It also needs to know the actual position of the column and the number of members if there is a merge column.

3. ILUSTRATION

Looking at figure 3 and table 2, and it needs to do the illustrations for the four algorithms. Algorithm 1, when read <tr> tag in the tag <table> ...</ table> then it will be known to the actual number of row is 7 (in the example table 2: *RsTotal* = 7) and the actual number of columns is 10 (*CsTotal* = 10) which obtained from the <td> tag in the first tag <tr>... <tr>.

In algorithm 2, to calculate the area of property obtained maximum value of row area of the property is 3 by looking the number of *rowspan* > 1 out of every tag <td> ...</ td> which read. Results of Algorithm 2 used by Algorithm 3 to define the property / attribute of the table 2 by first looking for the actual column position (*PosisiCol*) and the number of members if there is merge column (*jumAnggota*). Table 2 in the example has 10 attributes, namely: i). NO, ii). NAME + FIRST, iii). NAME + MIDDLE, iv). NAME + LAST, v). DATE OF + BIRTH + D, vi). DATE OF + BIRTH + M, vii). DATE OF + BIRTH + Y, viii). DATE OF DEATH + D, ix). DATE OF + DEATH + M, x). DATE OF + DEATH + Y.

From the results of Algorithm 3, the next step is to retrieve the content of data corresponding to an attribute that has been defined and the corresponding cell of the data. In Table 2 has cell as much as 4 rows x 10 columns. The results of table extraction will be stored in a database format.

4. CONCLUSION

Refer to above results of illustration, the algorithm can extract table in HTML format from Internet sources. Simple and complex format of tables especially for attribute have been able to construct to database with appropriate property.

Future work of the research can shift from single table in an URL address to many tables in an URL address. The main problem is to define which table that process to be extracted for the interoperability. For illustration, if there is a travel agent URL has three tables: price of airline, price of train and price of hotel, how to extract just table for price of train.

5. REFERENCES

[1] Embley, D.W, Tao, C. 2005. Automating the Extraction of Data from HTML Tables with Unknown Structure. *Data & Knowledge Engineering Vol. 54 Issue 1*, ISSN 0169023X, Elsevier

[2] Gultom, R.A.G, Fitri Sari, R, Budiarjo, B. 2011. Proposing the new Algorithm and Technique Development for Integrating Web Table Extraction and Building a Mashup. In *Journal of Computer Science 7 (2) : 129-142, ISSN 1549-3636*

[3] Hong, J.L., Fauzi, F., 2010. Tree Wrap-data Extraction Using Tree Matching Algorithm. *Majlesi Journal of Electrical Engineering vol 4 No. 2*

[4] Krupl, B, Herzog, M, Gatterbauer, W. 2005. Using Visual Cues for Extraction of Tabular Data from Arbitrary HTML Documents. *WWW Spesial Interest Tracks and Posters*, ISBN 1595930515, Association for Computing Machinery (ACM)

[5] Lerman, K., Knoblock, C., Minton, S., 2001. Automatic Data Extraction from Lists and Tables in Web Sources. In *Proceedings of Automatic Text Extraction and Mining workshop (ATEM-01),* IJCAI-01

[6] Lin, J., Wong, J., Nichols, J., Cyper, A. 2009. End-User Programming of Mashups with Vegemite. In *Proceedings of the 13th international conference on Intelligent user interfaces pp. 97-106*

Solving Shortest Path Problem Using Viral Systems

Dedy Suryadi
Parahyangan Catholic University
Ciumbuleuit 94
Bandung 40141, Indonesia
+62 22 2032700
dedy@unpar.ac.id

Marvin Antonie
Parahyangan Catholic University
Ciumbuleuit 94
Bandung 40141, Indonesia
+62 22 2032700

ABSTRACT
The shortest path problem is defined as a problem of finding the shortest path from a particular node to another, based on the given network connecting the nodes. This research proposes an algorithm using Viral Systems to solve the problem. Viral Systems applies the analogy of viruses infecting an organism. In Viral Systems, a solution is encoded into a genome. This research uses an indirect encoding, in which each node is given a priority value to be appended into the path while considering the network as well. The algorithm is implemented to 3 cases and the result is satisfactory. It is also found that the interactions of Viral Systems' parameters affect the solution quality significantly.

Keywords
Shortest path problem, Viral Systems, metaheuristic.

1. INTRODUCTION
The shortest path problem is one of the most extensively researched problems. The problem can be simply defined as finding the paths from a particular node to another that result in the shortest distance. In reality, the shortest path problem may be applied to the matters such vehicle routing in transportation system, traffic routing in communication networks, and path planning in robotic systems [1].

Some of the optimization methods known to solve shortest path problem are Dijkstra algorithm and dynamic programming [2]. Instead of using those well-known optimization methods, this research solves the problem using Viral Systems, which is a metaheuristic algorithm. Although metaheuristics never guarantee the achievement of an optimal solution, they have the advantage of obtaining a reasonably good solution within a relatively shorter time than optimization methods [3]. Therefore, metaheuristics are particularly useful when the problem is big but the available time to solve it is quite limited.

Viral Systems is a relatively new metaheuristic. Other metaheuristics that have been applied to solve the shortest path problem are GA (Genetic Algorithms) [4], ANN (Artificial Neural Networks) [5], and PSO (Particle Swarm Optimization) [1]. Viral Systems utilizes the analogy of viruses infecting the cells of an organism. The ending comes either when the viruses fully conquer the organism or, on the contrary, the organism successfully isolates the viruses. The algorithm is quite promising, as it performs better than the best-known metaheuristics for Steiner problem [6].

There will be five sections following this introduction. The next section introduces the Viral Systems algorithm. It is followed by the explanation of genome encoding specific for the shortest path problem. Later, the whole algorithm is presented and implemented into several experimental cases. Finally, some conclusions can be drawn from the experimental results and discussions.

2. VIRAL SYSTEMS
A virus consists of nucleic acid, such as DNA or RNA, and protein. The nucleic acid is contained inside a capsid generated by the protein. Thus, a virus may also be called as a nucleus-capsid. There are several types of viruses, such as Reovirus, Rhabdovirus, Picornavirus, etc. Each type of virus goes through different mechanism of replication. Each type applies different strategy to weaken its host cell as well.

Once a virus has infected a (bacteria) cell, which is called a phage, it may go through either lytic or lysogenic replication. In lytic replication, simply said, the virus makes the copies of itself inside the infected cell. After a particular number of copies has been created, they break the cell's border and infect close cells, which may be labeled as neighboring cells. An illustration for lytic replication is shown in Figure 1.

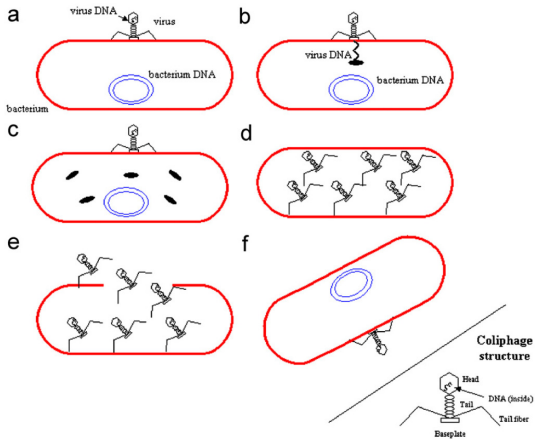


Figure 1. Lytic replication[6]

In lysogenic replication, however, a virus stays attached to the genome of the cell which it has infected. By attaching itself to the cell's genome, it has altered the cell's original genome as well. It remains that way until it is activated by a particular cause. Once activated, it would start the replication. An illustration for lysogenic replication is shown in Figure 2.

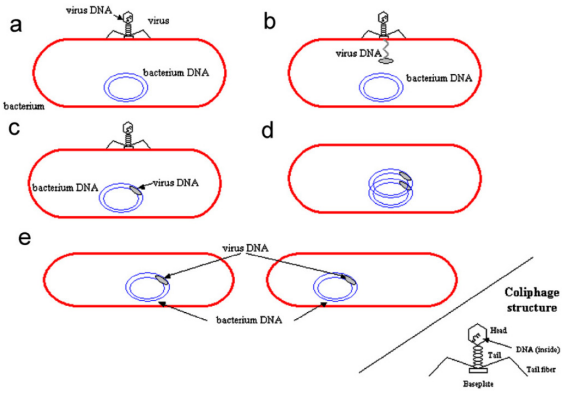


Figure 2. Lysogenic replication (right) [6]

In Viral Systems, the viruses are continuously looking to infect the cells with lower health. Nevertheless, a cell also has its defense mechanism to fight intruders, such as viruses. It can produce an antigene, which is an immune response to fight the virus. For those unable to produce an antigene, they are collected in a clinical picture. In computational term, a clinical picture collects the solutions on hand at a particular iteration. An illustration of a clinical picture is shown in Figure 3.

For each infected cell entering the clinical picture, initially it has NR status equals 0 and IT status equals 0 as well. NR is related to lytic replication, while IT is related to lysogenic replication. If the cell undergoes lytic replication, its NR status will be updated as many as the nucleus-capsids replicated. The number of nucleus-capsids replicated (Z) is governed by the following equation:

$$\mathrm{P}(Z=z)=\binom{\mathrm{LNR}}{z} p_r^z(1-p_r)^{\mathrm{LNR}-z}$$

where pr is the probability of replicating a nucleus-capsid, while LNR is the limit number of nucleus-capsids which allows the nucleus-capsids to break the cell's border.

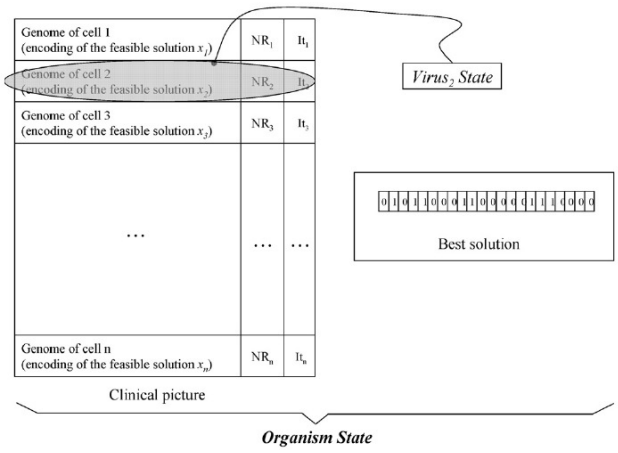


Figure 3. Clinical picture example [6]

The value of LNR is governed by the following equation:

$$\mathrm{LNR}_{\text{cell-}x}=\mathrm{LNR}^0\cdot\left(\frac{f(x)-f(\widehat{x})}{f(\widehat{x})}\right)$$

where LNR^0 is a constant multiplier, and $\widehat{x}$ is the cell producing the best-known objective function value (fitness) so far.

Each time a cell goes through lysogenic replication, its IT status is added by 1. Mutation occurs when a cell's IT status becomes larger than or equal to LIT. The LIT is computed by the following equation:

$$\mathrm{LIT}_{\text{cell-}x}=\mathrm{LIT}^0\left(\frac{f(x)-f(\widehat{x})}{f(\widehat{x})}\right)$$

where LIT^0 is a constant multiplier, and $\widehat{x}$ is the cell producing the best-known objective function value so far.

In computational term, the ending of the algorithm occurs under each of the two conditions. The first condition is the optimal solution has been obtained, or a solution whose quality is less than the determined gap with the optimal has been obtained. This is analogical to the condition when viruses successfully conquer the organism. The second one is the maximum number of iterations has been reached. This is analogical to the ability of the organism to finally isolate the viruses. This research applies the second condition, i.e. maximum number of iterations, as the ending criterion for the algorithm.

3. GENOME ENCODING FOR SHORTEST PATH PROBLEM

In order to explain the genome encoding, a simple network is presented in Figure 4 as an example to assist the explanation. The network consists of nine nodes and the objective is to find the shortest path from Node 1 to Node 9. The distance between any two nodes is presented as the number above the arc connecting the respective nodes.

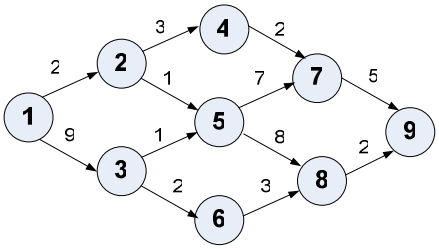


Figure 4. A network example

For a problem that involves n number of nodes, the genome consists of n genes as well. The first gene corresponds to the starting node, thus the value given for the gene is 1. For each of the remaining genes, an integer between 1 and 99 is taken randomly and assigned to that particular gene. These numbers do not directly represent the solution, i.e. the path chosen from the starting node to the ending node, but they serve as the priority value assigned to each node. An example of a genome for the network in Figure 4 is shown in Figure 5.

	2	3	4	5	6	7	8	9
1	25	34	11	22	25	3	12	44

Figure 5. A genome example

The genome example in Figure 5 may be translated into a solution by considering the priority values on the genome and the

relationship between nodes as depicted on the network in Figure 4. Since node 1 is the starting node, node 1 becomes the first node to be chosen. From node 1, there are two possible nodes to visit, i.e. nodes 2 and 3. Based on the genome, the second gene contains value 25 and the third gene contains value 34. Since the third gene contains a higher value, thus the next node chosen is node 3. From node 3, node 6 is chosen because it has higher priority value than node 5 –which is also possible to be visited from node 3. Afterward, node 8 must be chosen from node 6, as it is the solely possible node to visit. Finally, node 9 is chosen for the same reason. Therefore, the genome in Figure 5 may be translated to become a path of 1 – 3 – 6 – 8 – 9. The total length of this path is 16.

It has been described above that at the end of lytic replications, the replicated nucleus-capsids would infect the neighboring cells. The neighboring cells of a cell are assumed to have similar genome to the particular cell. The neighboring cells should be defined in a way such that the neighbors would not be either too many or too few. Too many neighbors may be good in terms of widening the searching space, but it certainly slows down the algorithm. Too few neighbors, on the other hand, limit the searching space, although it surely makes the algorithm faster.

Based on the consideration above, the neighboring cells of a cell are defined to be the ones having genome that contains the same values with that particular cell. However, the sequence of the values is not the same, instead there are two values are exchanged. The exchange does not apply to the first gene. In general, a genome with the length n has (n-2) neighboring cells. Consider a cell with the genome as shown in Figure 5, it has 7 neighbors as shown in Table 1.

Table 1. Neighboring cells

Neighbor	Exchanged Genes	Genome	Solution
1	2 ↔ 3	1-34-25-11-22-25-3-12-44	1-2-5-8-9
2	3 ↔ 4	1-25-11-34-22-25-3-12-44	1-2-4-7-9
3	4 ↔ 5	1-25-34-22-11-25-3-12-44	1-3-6-8-9
4	5 ↔ 6	1-25-34-11-25-22-3-12-44	1-3-5-8-9
5	6 ↔ 7	1-25-34-11-22-3-25-12-44	1-3-5-7-9
6	7 ↔ 8	1-25-34-11-22-25-12-3-44	1-3-6-8-9
7	8 ↔ 9	1-25-34-11-22-25-3-44-12	1-3-6-8-9

At the end of lysogenic replication, the cell's genome is mutated by the virus. Therefore, it is also required to set a procedure for mutating a genome. In this research, the mutation is done by selecting a gene randomly –excluding the first gene. The value of the selected gene is modified by exchanging the digits, i.e. the first digit becomes the second, and vice versa. The remaining genes stay exactly the same. For example, from the genome shown above, the third gene is selected. Its original value is 34. It is mutated to become 43. Thus, the genome after mutation is 1-25-43-11-22-25-3-12-44. If the original value consists of a single digit, for example 7, then it is considered to be "07", such that the mutation result becomes 70.

4. THE ALGORITHM

The Viral Systems for solving the shortest problems is generally described as follows.

1. Enter the inputs related to the shortest path problem, i.e. the number of nodes, the arcs connecting a node to the others, and the distance of each arc.
2. Enter the inputs related to Viral Systems, i.e. the constant multiplier for lytic replication (LNR^0), the constant multiplier for lysogenic replication (LIT^0), the size of clinical picture (POB), maximum number of iterations (ITER), the probability of undergoing lytic replication (plt), the probability of replicating a nucleus-capsid (pr), the probability of infecting neighboring cells (pi), and the probability of producing antigene (pan).
3. Create the initial clinical picture with the size of POB. Fill the clinical picture with as many cells as POB, avoid duplicated genomes. For each cell, create its genome and set its NR status equals 0, as well as its IT status. Decode the cell's genome and compute its objective function value.
4. Generate a random number to determine whether the cell is able to produce antigene or not, by comparing the number with pan. If a particular cell is able to produce antigene, it is deleted from clinical picture. Otherwise, it is not.
5. Record the best objective function value which has been obtained so far.
6. For each cell, generate a random number to determine whether it undergoes lytic replication or lysogenic replication, by comparing the number with plt. If the type of replication is lytic, go to Step 7. Otherwise, go to Step 12.
7. Update the cell's NR.
8. Compute the cell's LNR.
9. If NR < LNR, the cell's border is not broken yet. Go back to Step 6. Otherwise, the cell's border is broken and the nucleus-capsids starts infecting its neighboring cells.
10. Create the neighboring cells. For each neighboring cell, generate a random number to determine whether the cell is infected or not, by comparing the number to pi.
11. For each infected neighboring cell, generate a random number to determine whether it is able to produce antigene or not. If it is able to produce antigene, it is deleted. Otherwise, compute its objective function value and it is kept in a set containing infected neighboring cells. Go back to Step 6.
12. Update the cell's IT.
13. Compute the cell's LIT.
14. If IT < LIT, the cell is not mutated yet. Go back to Step 6. Otherwise, the cell is mutated. Compute the objective function value of the mutation result. It is in a set of mutation results. Go back to Step 6.
15. Update the best objective function value found so far, if necessary.
16. Create a set which is the union of infected neighboring cells set and mutation results set. Delete duplicates, if there is any. Check whether clinical picture has enough empty spaces to contain all cells in the set. If there are enough spaces, all cells in the set enter the clinical picture. Otherwise, delete the cell in the current clinical picture whose objective function value is the worst. It is repeated until there is enough space. If the set contains more members than POB, delete all cells in the current clinical picture and select randomly among the set to enter the clinical picture.

17. Check whether ITER has been reached. If it has been reached, the algorithm has been finished. Otherwise, go back to Step 4.

5. RESULTS and DISCUSSIONS

The algorithm is implemented to three cases. The first case involves 9 nodes and the network is shown in Figure 4. The optimal solution is known to have objective function value 12. The algorithm obtains the optimal solution with the shortest path 1-2-4-7-9.

The second case involves 20 nodes [1]. The optimal solution is known to have objective function value 142. The optimal solution is obtained by the algorithm with the shortest path 1-3-8-14-20.

The third case involves 50 nodes and its network is shown in Figure 6. The best solution found for the third case has objective function value 185, with the shortest path 1-5-10-17-24-37-46-49-50.

The third case is used for Viral Systems' parameters testing. The tested parameters are:

- plt: 0.1 and 0.9
- pi: 0.1 and 0.9
- pr: 0.1 and 0.9
- LNR^0: 5 and 25
- LIT^0 : 5 and 25

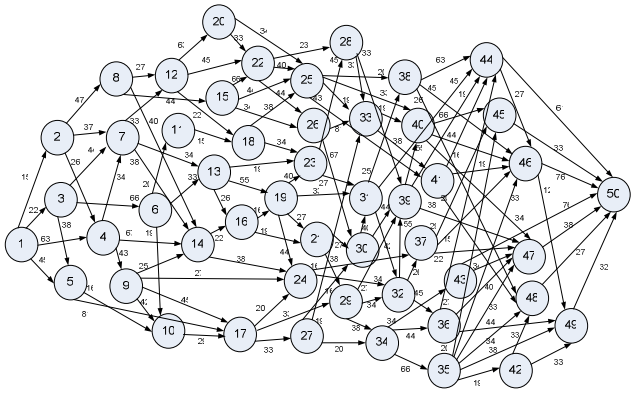


Figure 6. Network for the third case

Therefore, there are 32 combinations of those parameters. Each combination is run for 15 replications. The parameters POB, ITER, and pan are not tested because their effects are deemed to be relatively known. The larger value of POB and ITER should give the better solution as there are more chances to explore the solution space. For pan, the value is determined by the inequality:

$$p_{an} > \frac{n \cdot \pi_{LNR}(p_i \cdot |\overline{V(x)}| - 1)}{n \cdot \pi_{LNR}(p_i \cdot |\overline{V(x)}| - 1) + n}$$

where π_{LNR} is the steady state probability of having LNR nucleus-capsids and V(x) is the average number of neighboring cells.

Based on the statistical test ANOVA ($\alpha = 5\%$), the parameters that affect the objective function value significantly are:

- The interaction between plt and pr
- The interaction between plt and LIT^0
- The interaction between plt and LNR^0
- The interaction between plt, pi, and pr

There are quite some parameters to be determined in this algorithm, and apparently their interactions significantly affect the quality of solutions. Although not tested in this experiment, pan itself is an interesting parameter. Using Equation (3), sometimes pan is required to be so high (roughly more than 0.9). Such extremely high pan obviously causes the cell to be very able to produce antigene. In other word, the obtained solutions have very small chance to either explore its neighborhood or undergo mutation. Before any of those occurs, the cell has been able to be free from the infection and thus leaving clinical picture. Therefore, it is worthwhile to explore more about the pan, particularly about releasing the inequality constraint.

The first experimental case is taken from a previous research which applies Genetic Algorithms to solve the TSP [10]. The case involves 10 cities. The distance between cities is shown in Table 1.

6. CONCLUSIONS

The Viral Systems for solving the shortest path problem has been proposed in this research. The preliminary experiments show that the algorithm is certainly able to achieve the optimal solution for the first two cases. For the third case, the objective function value achieved is also relatively consistent.

More experimental cases are suggested for the future research in order to examine more thoroughly the effect of each parameter. It is also suggested to consider abandoning the inequality for pan.

7. REFERENCES

[1] Mohemmed, A.W., Sahoo, N.C., Geok, T.K., 2008, "Solving Shortest Path Problem Using Particle Swarm Optimization," Applied Soft Computing, Volume 8, Issue 4, 1643-1653.

[2] Winston, W.L., 2004. "Operations Research: Applications and Algorithms", 4th Edition, Duxbury Press, Belmont.

[3] Reeves, C.R., 1995, "Modern Heuristic Techniques for Combinatorial Problems,", McGraw-Hill, United Kingdom.

[4] Gen, M., Cheng, R., Wang, D., 1997, "Genetic Algorithms for Solving Shortest Path Problems," Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC '97), Indianapolis, USA.

[5] Wang, J., 1996, "A Recurrent Neural Network for Solving the Shortest Path Problem," IEEE Transactions on Circuits and Systems Part I: Fundamental Theory and Applications, Volume 43, Issue 6.

[6] Cortés, P., Garcia, J. M., Munuzuri, J., Onieva, L., 2008, "Viral Systems: A New Bio-Inspired Optimation Approach," Computers & Operations Research 35, 2840–2860.

Fitness Evaluation of Multi-Element Genetic Algorithm for Traffic Signal Parameters Optimization

I Gede Pasek Suta Wijaya
Electrical Eng. Dept, Mataram Univ.
Jl. Majapahit 62 Mataram NTB
+62-370-636126
gpsutawijaya@te.ftunram.ac.id

Keiichi Uchimura
GSST Kumamoto University
2-39-1 Kurokami JAPAN
+81-96-3638
uchimura@cs.kumamoto-u.ac.jp

Gou Koutaki
GSST Kumamoto University
2-39-1 Kurokami JAPAN
+81-96-3638
koutaki@cs.kumamoto-u.ac.jp

Shinichi Isigaki,
Network Eng. Application Lab. LTD
Hakata 1-4-4 Fukuoka JAPAN
ishigaki@nalab.jp

Hiroshi Sugitani
Network Eng. Application Lab. LTD
Hakata 1-4-4 Fukuoka JAPAN
sugitani@nalab.jp

ABSTRACT
This paper presents evaluation of several fitness formulations for optimizing traffic light signal parameters using Multi-Element Genetic Algorithms (ME-GA). In this evaluation, the ME-GA is integrated with the AIMSUN 6.1 simulator for finding out the best traffic signal control parameters correspond to the best fitness formula. To know the performance of the proposed methods, several experiments were carried out using simple network model and real network model (*Oee-Toroku Kumamoto City* Road Network Model). The experimental results show that the fitness model IV could provide the most optimum traffic signal parameters which is shown by higher percentage of vehicle flow and less vehicle delay time than those of the real word.

Keywords
Traffic signal control, optimization, transport system, intelligence system, and GA.

1. INTRODUCTION
The modern city need a fast, safe, and inexpensive transportation system in order to minimize the effect of the traffic congestion on economics and environmental. One of the most methods dealing with this problem is traffic signal control optimization. By this optimization, large vehicle flow, small number of vehicle stop, short delay time, and maximum throughput can be achieved.

A multi-Element Genetics Algorithms (ME-GA) which is an concept of GA-based algorithm for optimizing traffic signal control containing of multi traffic light signal parameters has been proposed to solve this problem. However, this solution still does not work well for road network containing extreme networks state. In this paper, we investigate some fitness formulations to get the most optimum traffic signal parameters. In this case, the ME-GA is integrated with the AIMSUN 6.1 simulator through the API module for simulation. Several experiments were done using simple network model and real network model (*Oee Toroku Kumamoto City* Road Network Model) which main aim whether the proposed method can provide high percentage of vehicle flow, maximum travel distance and short delay time for these models.

2. THE PREVIOUS WORKS
The traffic signal optimization has been considered as one of traffic jam solution. This solution is chosen because it does not need to change many resource of the transportation system. This solution just need to readjust the signal parameters of traffics light of correspond road network.

Several methods regarding to optimization of the traffic signal control have been proposed as presented in Refs. [1-7]. Those methods can be categorized as an optimization based on Genetic Algorithm approach[1-2], fuzzy logic approach[4-5], on stochastic approach[6], and dynamic approach [7], respectively. The GA and fuzzy logic-based algorithm called as intelligence system based traffic signal optimization have been proved to provide an optimum signal parameter on their network models. The GA on the previous methods mostly implemented for optimizing cycle and offset parameters and they did not implement on real network model.

In addition, the Ref. [8-9] also proposed the multi element genetic algorithms for optimization of traffic signal control; however, it did not reach 100% of vehicle flow percentage. In addition, the Ref. [10] also did not implemented for optimizing the real network model.

This paper presents evaluation of several fitness formulations for finding out the most optimum traffic signal parameters correspond to the best fitness formula for Multi-Element Genetic Algorithms (ME-GA).

3. THE PROPOSED METHODS
3.1 ME-GA
Suppose we have road network, as shown in Fig. 1. From this road network, we design the ME-GA, which can be implemented for optimizing traffic signal control parameters, using encoding process. The encoding process wants to extract a set of traffic signal control parameters, which have to be optimized. Each node of road network (see Fig. 1) consists of four signal parameters: firstly, offset representing the time coordination between traffic light; secondly, cycles representing the total time of traffic light

starting from Green then returning to Green; thirdly, Yellow and All Red (AR) representing the time for Yellow and All Red signal and finally, main and sub split which is Green time percentage of main road and sub road, respectively[10]. Usually, the Yellow and AR is defined as a constant value.

By encoding process, a simple road network (see Fig 2) which consists of four junctions can be encoded into some element of GA such a population (Pop), as shown in the Table 1. Each junction/node contains of a chromosome consisting of offset, cycle, and split parameters.

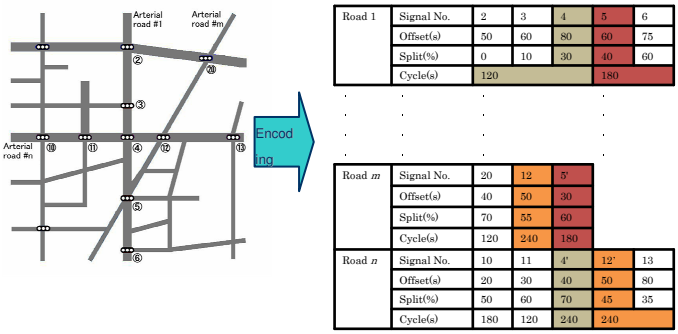


Figure 1. Network encoding into multi signal parameters[8,9,10]

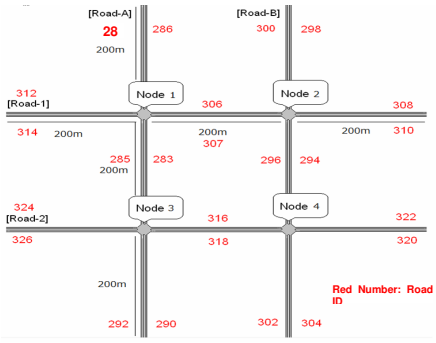


Figure 2. A Simple Network Model

Table 1. Encoded population of the simple road network model

Pop [*i*], *i*=0, 1, 2, …, *N*-1

Node	Offset(s)	Cycle(s)	SplitMain(%)	SplitSub(%)	
01	0	130	60	40	← Chromosom [0]
02	58	140	40	60	← Chromosom [1]
03	25	120	65	35	
04	53	150	40	60	← Chromosom [*M*-1]

Next, from the encoded traffic control signal parameters, the ME-GA based the optimization algorithms is created which is shown by flow chart in Fig. 3. The optimization process starts from initializing the ME-GA for *N* populations. Then each population is sent to the simulator for setting the traffic signal control and then simulates the network to get the vehicle flow, total distance (t_{Dist}), and delay time (t_{Delay}) of the passing vehicle in the network. The vehicle flow is determined from vehicle gone out (V_{go}) which means the total vehicles that have gone out from the network, vehicle in the network (V_{in}) which means the total vehicles that still leave in the network, and vehicle wait out (V_{wo}) which means the total vehicles which are waiting for entering into the network.

From the V_{go}, V_{in}, V_{wo}, t_{Dist}, and t_{Delay} output parameters, each population is evaluated by determining the fitness. In this case, we investigate four type of finesses called as fitness I, II, III, and IV.

Next from the obtained fitness, crossover and mutation are done for obtaining the new populations. All of these processes are repeatedly performed until reaching the maximum generation. In addition, the ellists algorithm is also included in the recombination of the populations. Finally, the best population giving the smallest fitness is determined from the last generation.

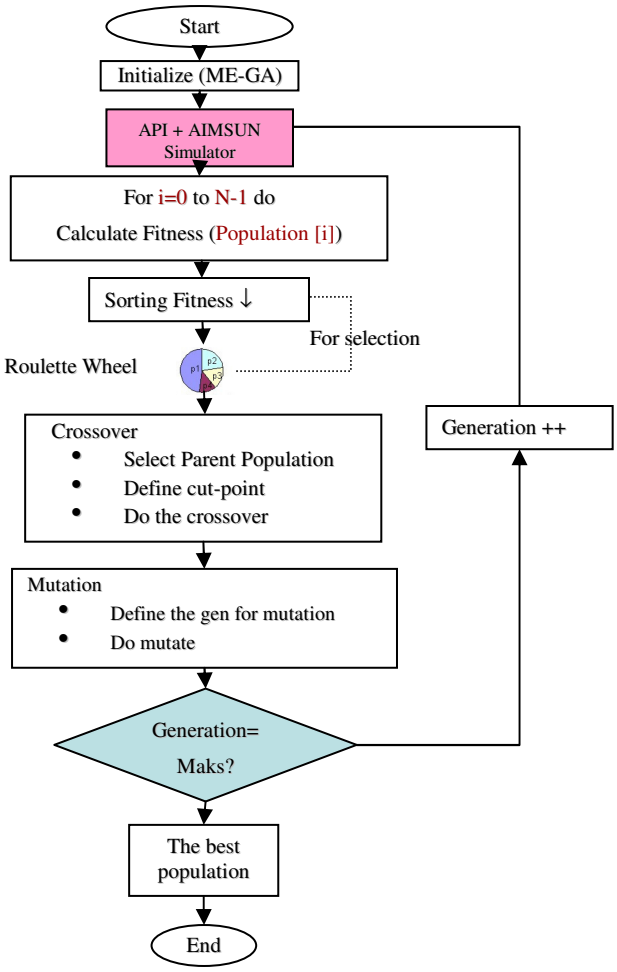


Figure 3. Flow chart of ME-GA[10]

3.2 Fitness Formulations

The Fitness I is determined by just considering the t_{Delay} output and the input cycles parameters. The aim of this fitness is try to find out the best traffic signal parameters, which can give the short t_{Delay}, large V_{go}, small V_{in} and V_{wo}, and short cycles time. The defined Fitness I can be expressed by the following equations [8]

$$F_E = \exp\left(\frac{t_{Delay}}{C}\right), \quad (1)$$

$$F_S = \exp\left(\frac{Cycles - C_a}{C_b}\right), \text{ and} \quad (2)$$

$$F_I = F_E \times F_S \quad (3)$$

where C, C_a, and C_b are constant values which are given by 50, 40, and 100, respectively. In addition, t_{Delay} is a delay time output from simulator in second.

The Fitness II is defined using the division of the t_{Delay} and V_{go}. From trial and error of simulation results, we found that t_{Delay} depends on number of vehicles flow to the network (the more vehicles flow are the more the delay time is). Therefore, we take the division of the delay time and V_{go} in order to get normalize value of the t_{Delay}. The main aim is to get the smaller t_{Delay}, the larger V_{go}, and less V_{in} and V_{wo} for the simulation than those of the Fitness I.

$$t_D = \frac{t_{Delay}}{V_{go}} \quad (4)$$

Then from the t_D, the $F_E^{'}$ is defined as follows,

$$F_E^{'} = \exp(t_D)\,. \quad (5)$$

Next, in order to synchronize the t_{Delay} and the vehicles flow, the next equation is defined.

$$F_F^{'} = \exp\left(1 - \frac{V_{go}}{V_{go} + V_{wo} + V_{in}}\right) \quad (6)$$

Finally, the fitness II is defined as :

$$F_{II} = F_E^{'} \times F_F^{'} \quad (7)$$

The fitness III is defined by considering the V_{wo}, V_{in}, t_{Delay}, and Q_{max}. The main aim to find out the signal parameters which can provide the smallest values of V_{wo}, V_{in}, t_{Delay}, and Q_{max} as much as possible.

$$F_{III} = \exp(\frac{V_{wo}}{C_{wo}}) + \exp(\frac{V_{in}}{C_{in}}) + \exp(\frac{t_D}{C_{tD}}) + \exp(\frac{Q_{\max}}{C_Q}) \quad (8)$$

where C_{wo}, C_{in}, C_{tD}, C_Q are constant values of vehicle wait out, vehicle in, delay time and maximum queue which are given as follows: C_{wo} = 100, C_{in} =500, C_{tD} =500, and C_Q =1000

Finally, the fitness IV is defined by considering not only V_{wo}, V_{in} but also the normalized t_{Delay}. In addition, because of the value of Q_{max} is almost the same as V_{wo}, therefore the Q_{max} is removed from the fitness.

$$V_f = \exp(\frac{V_{wo}}{C_{wo}}) + \exp(\frac{V_{in}}{C_{in}}) + \exp(\frac{t_D^{'}}{C_{tD}}) \quad (9)$$

where $t_D^{'} = t_D \Big/ tot_{Tr}^{D}$, t_D= delay time and tot_{TR}^{D} =total travel distance. In this case, the constant values (C_{wo}, C_{in}, C_{tD}) are setup as the same as those of the Fitness III.

The integration of ME-GA with the AIMSUN 6.1 simulator via the API that is provided by this simulator is shown in Fig 4. In this case, the ME-GA is inserted into API module and the API is built into dynamic library (dll). On the simulation, firstly, the simulator for creating the random traffic signal control parameters executes the API-dll; next, the simulator set the traffic light signals of the tested road network using the created signals and run the simulation; finally, the simulator recalls the API-dll to evaluate the output of simulator. These processes are repeatedly done until reaching the maximum generation.

4. EXPERIMENTS AND RESULTS DISCUSSION

4.1 Experiment Setup

The experiment setup was prepared before running the simulation, which consists of creating road network model, setting-up GA, and setting up the AIMSUN 6.1 parameters, as follows:

1. Preparing two road network models: simple network model (see Fig. 2) and *Oee Toroku* network model as real traffic signal problem, as shown in Fig. 5.
2. Setting up the GA and AIMSUN 6.1 using data as given in Table 2. The lane width and the vehicle size were defined just for the simple network model while the real network model was setup as the same as the real conditions of *Oee Toroku* site.

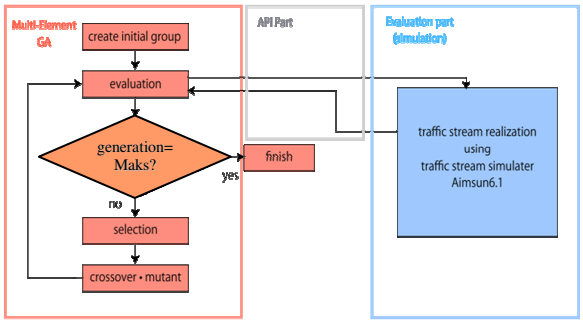


Figure 4. The Intgeration of ME-GA with the AIMSUN 6.1 Simulator

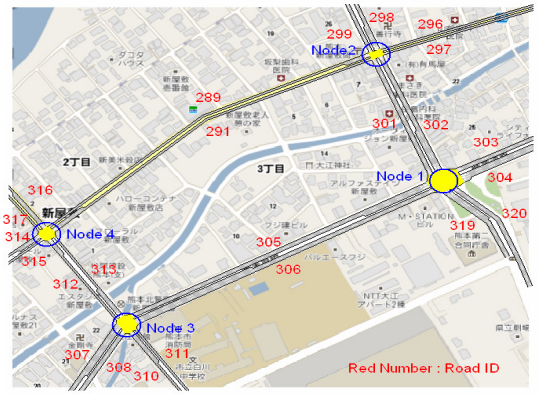


Figure 5. The *Ooe Toroku* road network model for experiments[10]

4.2 The Fitness Evaluations of ME-GA on Simple Road Network

In this case, we wanted to find out the best fitness formulation for ME-GA by evaluating four fitness formulations on simple network model using three kinds of network states. In addition, the following constraints were used for the evaluations:

- Offset : Min =0, Max = 120 and it incremental, Δ_{offset}=1,
- Cycle : Min =60, Max = 180 and it incremental, Δ_{Cycle}=5,
- Split : Min =10, Max = 90 and it incremental, Δ_{Split}=5, and
- Signal timing (t) is determined from green time (gT) only.
- Experiments were done using 5 minutes warming up.

Table 2. The GA and AIMSUN 6.1 data setup

GA		AIMSUN 6.1	
Generation number	25	Lane width	3 m
Individual number	50	Vehicle width	2 m
Crossover rate	80 %	Branch	Only straight
Mutant rate	5 %	Speed	50 km/h
Crossover method	2 point crossing	Number of vehicles	4800 vehicles/h
Selection method	Roulette selection	Acceleration	30 m/s^2
Evaluation formula	3 Variations of fitness	Deceleration	40 m/s^2

A. The evaluation on the network state 1

The network state 1 consists of the straight way and turn right configuration, which the state data setup are shown in the Table 3. The aim of this simulation was to know whether the proposed fitness for ME-GA could provide best traffic parameters or provide network crash for this state.

Table 3. Network state 1 data setup for simple network.

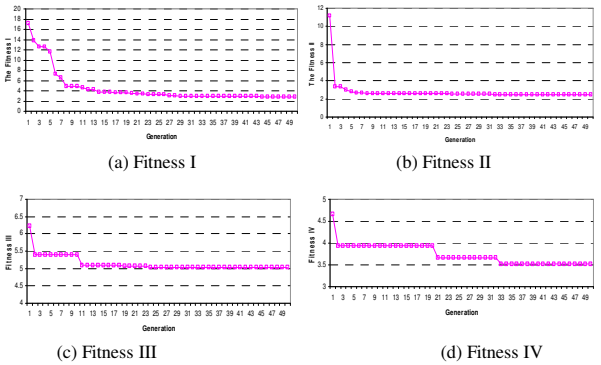
(a) Vehicle flow per hour (VF)

Road ID	286	292	298	302	310	312	320	324	**Total**
VF	800	0	400	0	0	400	0	800	**2400**

(b) Turning percentage of VF

No	Turning	T.P.	No	Turning	T.P.	No	Turning	T.P.	No	Turning	T.P.
1	283/290	50	9	294/304	50	17	306/308	50	25	316/322	50
2	283/316	50	10	294/322	50	18	306/300	50	26	316/296	50
3	285/288	50	11	296/300	50	19	307/314	50	27	318/326	50
4	285/314	50	12	296/307	50	20	307/283	50	28	318/290	50
5	286/283	50	13	298/294	50	21	310/307	50	29	320/318	50
6	286/306	50	14	298/308	50	22	310/294	50	30	320/304	50
7	292/285	50	15	302/296	50	23	312/306	50	31	324/316	50
8	292/326	50	16	302/318	50	24	312/288	50	32	324/285	50

Note: 283/290 the turning percentage (T.P.) from rod ID 283 to 290 is 50%.



(a) Fitness I (b) Fitness II

(c) Fitness III (d) Fitness IV

Figure 6. ME-GA fitness on the network state 1

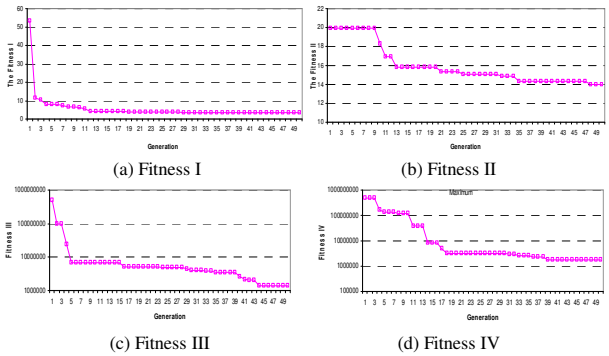
From the simulation results, we can see that all fitness formulations of ME-GA can provide the best traffic signal parameters, which is shown by good finesses graph outputs, as shown in Fig. 6. In addition, from the best signal parameters of fitness I ~ IV (see Table 4), the simulator give high V_{go} and less V_{in}, V_{wo}, and t_{Delay} as shown in experimental data summary (Table 9). It means all the fitness formulations of ME-GA tend to provide good solution for the tested network state of simple road network model. It can be achieved because this state contains of small vehicle flow 2400 veh/hours.

Table 4. The best signal parameters of the Fitness I, II, III, and IV on the network state 1

Junction	Cycle (s)	Split (%)	Offset (s)	Cycle (s)	Split (%)	Offset (s)
	Fitness I			Fitness II		
Node 1	50	30:70	50	180	65:35	106
Node 2	50	55:45	6	80	45:55	1
Node 3	50	60:40	58	60	60:40	1
Node 4	50	70:30	14	120	30:70	28
Junction	Fitness III			Fitness IV		
Node 1	175	70:30	-24	150	70:30	101
Node 2	145	45:55	-68	130	55:45	55
Node 3	95	50:50	-17	90	50:50	108
Node 4	140	40:70	-60	120	50:50	25

Table 5. The VF of network state 2 for simple network model

Road ID	286	292	298	302	310	312	320	324	**Total**
VF	800	400	400	800	800	400	400	800	**4800**



(a) Fitness I (b) Fitness II

(c) Fitness III (d) Fitness IV

Figure 7. ME-GA fitness on the network state 2

B. The evaluation on the network state 2

The network state 2 is the same as the network state 1 except for the number of VF of the network that is increased twice of that of network state 1, as shown in the Table 5.

From the fitness graph point of view of the evaluation on this state (See Fig. 7), all the fitness formulations for ME-GA also tends to provide good solution. However, from V_{go}, V_{in}, V_{wo}, and t_{Delay} point of view, only the best signal parameters (see Table 6) of the Fitness I have higher difference between the VF and V_{go} (Δ) than those of Fitness II, III, and IV, which are ideally zero, as shown in Table 9. It means there is no guarantee for the fitness formula which its fitness graph looks to be convergence, can give good V_{go}, V_{in}, V_{wo}, Δ, and t_{Delay} . Therefore, based on V_{go}, V_{in}, V_{wo}, Δ, and t_{Delay} parameters, the Fitness III tends to provide better performance than those of the Fitness I, II, and IV which is shown by smaller Δ and reasonable enough the delay time.

C. The evaluation on the network state 3

The network state 3 consists of the straightway and turn left con-

Table 6. The best signal parameters of the Fitness I, II, III, and IV on the network state 2

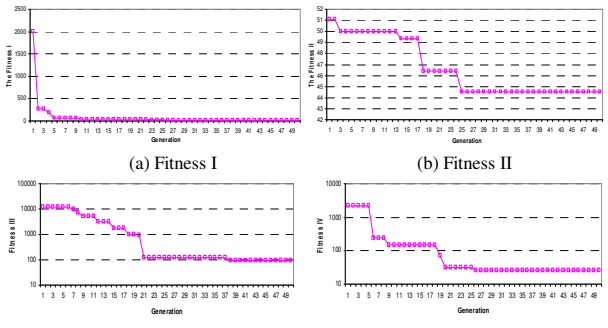
Junction	Cycle (s)	Split (%)	Offset (s)	Cycle (s)	Split (%)	Offset (s)
	Fitness I			Fitness II		
Node 1	50	35:65	71	165	55:45	112
Node 2	50	60:40	168	155	45:55	18
Node 3	50	65:35	2	85	60:40	3
Node 4	50	30:70	91	165	50:50	89
Junction	Fitness III			Fitness IV		
Node 1	135	55:45	27	165	50:50	55
Node 2	170	50:50	93	100	55:45	86
Node 3	135	55:45	115	155	50:50	110
Node 4	155	50:50	98	115	50:50	91

figuration, which is presented in Table 7. This state has the same VF as that of network state 2.

The same as the result for the networks state 2, the graph of the all Fitness evaluations of ME-GA (See Fig. 8) looks to be convergence for this network state. However, from V_{go}, V_{in}, V_{wo}, and t_{Delay} point of view of the best signal parameters (Table 8), the Fitness I also has high difference between the *VF* and V_{go} (Δ), as presented in Table 9. It also means, the Fitness I does not provide optimum signal parameters even though it delay time is almost 4 time less than that of the Fitness II. Therefore, based on V_{go}, V_{in}, V_{wo}, Δ, and t_{Delay} parameters, the Fitness IV tends to provide better performance than those of the Fitness I, II, and III which is shown by smaller Δ and reasonable short enough the delay time. Over all, the Fitness II~IV tends to give better performance than that of the Fitness I, which is shown by smaller Δ and short enough t_{Delay}.

Table 7. Turning Percentage of network state 2 for simple network model

No	Turning	T.P.	No	Turning	T.P.	No	Turning	T.P.	No	Turning	T.P.
1	283/290	50	9	294/304	50	17	306/308	50	25	316/322	50
2	283/326	50	10	294/318	50	18	306/294	50	26	316/304	50
3	285/288	50	11	296/300	50	19	307/314	50	27	318/326	50
4	285/306	50	12	296/308	50	20	307/288	50	28	318/285	50
5	286/283	50	13	298/294	50	21	310/307	50	29	320/318	50
6	286/314	50	14	298/307	50	22	310/300	50	30	320/296	50
7	292/285	50	15	302/296	50	23	312/306	50	31	324/316	50
8	292/316	50	16	302/322	50	24	312/283	50	32	324/290	50



(a) Fitness I (b) Fitness II

(c) Fitness III (d) Fitness IV

Figure 8. ME-GA fitness on the network state 3

Based on this simulation result summary, as shown in the Table 9, we can conclude that the Fitness IV always provides better traffic signal parameters shown by higher Δ parameter and reasonable short delay time for three tested network states. Therefore, for the next test on the real road network model, we just consider the Fitness IV and II as comparison to confirm whether the Fitness IV could find out the best traffic signal parameters, which can provide the most optimum V_{go}, V_{in}, V_{wo}, and t_{Delay} parameters, as done on simple road network model.

Table 8. The best signal parameters of the Fitness I, II, III, and IV on the network state 3

Junction	Cycle (s)	Split (%)	Offset (s)	Cycle (s)	Split (%)	Offset (s)
	Fitness I			Fitness II		
Node 1	140	60:40	52	140	50:50	50
Node 2	50	40:60	45	160	50:50	25
Node 3	90	35:65	58	140	50:50	48
Node 4	120	90:10	-41	105	55:45	68
Junction	Fitness III			Fitness IV		
Node 1	125	45:55	28	125	45:55	16
Node 2	130	50:50	112	175	40:60	26
Node 3	155	45:55	54	150	50:50	85
Node 4	175	50:50	41	130	50:50	14

Table 9. Experimental results summary of fitness evaluations of simple road network

No	Pattern	VF	Fitness	V_{go}	V_{in}	V_{wo}	$\Delta=VF-V_{go}$	Delay Time
1	State 1	2400	I	2219	77	168	181	117.7884
			II	2301	89	83	99	146.0541
			III	2321	132	18	79	189.3953
			IV	**2341**	**125**	**17**	**59**	**183.7815**
2	State 2	4800	I	1061	453	3516	3739	111.6548
			II	3186	347	1510	1614	450.2981
			III	3301	318	1416	1499	430.3488
			IV	**3269**	**336**	**1422**	**1531**	**470.9226**
3	State 3	4800	I	2020	351	2547	2780	57.798
			II	4320	204	441	480	188.409
			III	4430	278	263	370	197.464
			IV	**4431**	**231**	**299**	**369**	**183.7815**

4.3 The Fitness Evaluation of ME-GA on the Real Network Model

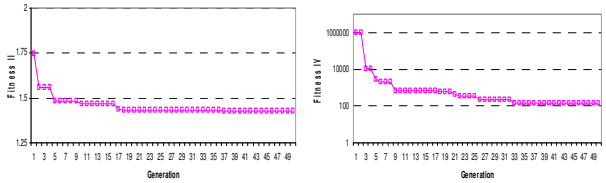
In this case, the vehicle flow of real road network state was setup using real data, as summarized in the Table 10. This data were captured manually by came to the *Oee Toroku* site, and counted the vehicles flow and turning vehicles at peaks sessions (8:00 AM to 9:00 AM). The peak traffic in this site always happens from 8:00 AM to 9:30 AM and from 5:00 PM to 6:30 PM. All of the vehicles sources were counted at the same day and the same time. In addition, the road network also considered the pedestrian flow, which were distributed in to four junctions/nodes by about 636, 1386, 415, and 860 people crossing on the junction/node 1, 2, 3 and 4 (see Fig. 5), respectively. Furthermore, the additional network properties such as road width and length and capacity were setup based on the real condition of the *Oee Toroku* site.

In this experiment, we evaluated two fitness (Fitness II and Fitness IV). The Fitness I and III were not considered because the fitness I was the worst one and the Fitness III almost had the same performance as that of the fitness IV. The main aim of this experiment is to know whether the proposed fitness could give the optimum traffic signal parameters, which could solve the traffic jam of the *Ooe Toroku* network problem at the peak sessions. From the fitness graph point of view, as shown in Fig. 9, the Fitness II and IV for ME-GA tends to achieve convergence and give the best signal parameter for this network model in the 50-th generation, as shown in Table 11. When the best signal parameters are implemented on the real network, they give smaller difference between the *VF* and V_{go} (Δ) and shorter vehicle delay time than those of the original signal parameters, as shown in Table 12. In addition, the best signal parameters of ME-GA for

Ooe Toroku network can improve the original percentage of vehicle flow by about 10.17% from 71.0% to 81.17% for fitness model II and 11.63% from 71.0% to 82.63% for fitness model IV.

Table 10. Vehicle flow of the *ooe toroku* network.

No	Road ID	Flow /hour		
		Car	*Bus*	*Truck*
1	297	486	8	24
2	298	586	18	28
3	304	1594	24	64
4	307	1122	30	64
5	310	318	0	12
6	314	164	0	6
7	316	432	2	30
8	319	456	12	30
Total		**5158**	**94**	**258**



(a) Fitness II (b) Fitness IV

Figure 9. ME-GA Fitness on the Oee Toroku network

Table 11. The best signal parameters of the Fitness II and IV on the *Ooe Toroku* Network

Junction	Fitness II			Fitness IV		
	Cycle (s)	Split (%)	Offset (s)	Cycle (s)	Split (%)	Offset (s)
Node 1	140	55:45	55	110	60:40	39
Node 2	140	50:50	50	110	55:45	12
Node 3	160	45:55	45	130	70:30	49
Node 4	155	60:40	60	130	70:30	86

Table 12. Simulation result summary on *Ooe Toroku* Network

Cases	Vehicle	VF	V_{go}	V_{in}	V_{wo}	Δ	Delay Time	tot_{TR}^{D} (km)	$F_F(\%)$
Real	Bus	94	81	20	32	13	NA	**NA**	**71.0**
	Car	5158	3085	1132	1340	2073			
	Truck	258	159	55	65	99			
	Pedestrian	3708	3239	34	0	469			
	Total	**9218**	**6564**	**1241**	**1437**	**2654**	**NA**		
Fitness II	Bus	94	72	9	16	22	940.227	**5237.92**	**81.17**
	Car	5158	3461	600	908	**1697**			
	Truck	258	210	36	56	48			
	Pedestrian	3708	3650	84	6	58			
	Total	**9218**	**7393**	**729**	**986**	**1825**	**0.251***		
Fitness IV	Bus	94	86	12	5	8	913.987	**5790.43**	**82.63**
	Car	5158	3784	894	381	**1374**			
	Truck	258	254	37	23	4			
	Pedestrian	3708	3453	162	79	255			
	Total	**9218**	**7577**	**1105**	**488**	**1641**	**0.222***		

* The Delay Time divided by V_{go} and $F_F=(1-F_F')*100$

These experimental results confirm and inline with the previous results on simple road networks, which shows that the Fitness IV, always provide the best signal parameters. It means these experimental results prove that the proposed Fitness IV of ME-GA can find out the best solution for traffic jam and give significant improvement for the tested road networks.

5. CONCLUSIONS AND FUTURE WORKS

The proposed fitness evaluations of multi element GA (ME-GA) have been successfully implemented for optimizing the traffic signal parameters. The best fitness IV is the best fitness model, which could provide the best solution for traffic jam of tested road networks. The small difference between the *VF* and V_{go} and short enough vehicle delay time for both simple network model and real network model problem can be given by the best signal parameters obtained by the Fitness IV for ME-GA.

In future, we will try to find out the best combination of population and generations by doing more experiments on the real road networks. In order to decrease the simulation time sub networks will be employed which can provide almost the same performances as those of without non-sub network model.

6. ACKNOWLEDGMENTS

I would like to send my appreciation and my great thank to the Image Media Laboratory of Kumamoto University and Network Engineering Application Laboratory LTD for supporting the research under SCOPE project scheme.

7. REFERENCES

[1] Singh, L., Tripathi, S. and Arora, H. 2009. "The Optimization for Traffic Signal Control Using Genetic Algorithm", International Journal of Recent Trends In engineering, vol. 2, no. 2, pp. 4-6.

[2] Ceylan, H. and Bell, M.G.H. 2004. "Traffic Signal Timing Optimization Based on Genetic Algorithm Approach, Including Drivers Routing", Transportation Research Part B, vol. 38, pp. 329-342.

[3] Takahashi, S., Kazama, H., Fujikura, T. and Nakamura, H. 2004. "Adaptive Search of an Optimal Offset for the Fluctuation of Traffic Flow using Genetic Algorithm", IEEJ Trans. on Industry Applications, vol. 123, Issue 3, pp. 204-210.

[4] Chou, C.H. and Teng, J.C. 2002. "A Fuzzy Logic Controller for Traffic Junction Signals", Information Sciences, vol. 143, pp. 73–97.

[5] Rahman, S.Y. and Ratrout, N.T. 2009. "Review of the Fuzzy Logic Based Approach in Traffic Signal Control: Prospects in Saudi Arabia", Transpn Sys Eng & IT, vol. 9, no. 5, pp. 58-70.

[6] X.H. Yu and Recker, W.W. 2006. "Stochastic adaptive control model for traffic signal systems", Transportation Research Part C, vol. 14, pp. 263–282.

[7] H.K. Lo, E. Chang, and Y.C. Chan, "Dynamic Network Traffic Control", Transportation Research Part A , vol. 35, pp. 721-744, 2001.

[8] Nishihara, T., Matsumura, N. Kanamaru, K., Wijaya, IGPS., Koutaki, G., Uchimura, K., Sugitami, H. and Ishigaki, S. 2011. "Construction of Dynamics Control System for Traffic Cogestiom Using Multi-Element GA", The 6th International Student Conference on Advanced Science and Technologhy, Jinan, China, pp. 139-140.

[9] Kanamaru, K., Matsumura, N., Wijaya, IGPS., Uchimura, K., Koutaki, G., Ishigaki, S. and Sugitami, H. 2011. "Construction of Software Framework for Simulation using Traffic Signal Simulator and Its Consideration", IPSJ SIG Technical Report, Fukuoka (CD-ROM).

[10] Wijaya, I GPS., Uchimura, K., Koutaki, G., Nishihara, T., Matsumoto, S., Ishigaki, S. and Sugitami, H. 2012. "Traffic Signal Control Parameters Optimization Using Multi-Element Genetic Algorithm", 2nd Annual International Conference on

Control, Automation and Robotics (CAR 2012), Bangkok Thailand, March 2012, (CD-ROM).

Efficient Sequential Access Method of Fingerprint Identification

G. Indrawan
Data & Software Engineering Research Division - STEI, ITB
Jl. Ganeshano. 10 Bandung, West Java, Indonesia, Ph. +62 22 2502260, Fax +62 22 2534222
gdindrawan@gmail.com

B. Sitohang
Data & Software Engineering Research Division - STEI, ITB
Jl. Ganeshano. 10 Bandung, West Java, Indonesia, Ph. +62 22 2502260, Fax +62 22 2534222
benhard@stei.itb.ac.id

S. Akbar
Data & Software Engineering Research Division - STEI, ITB
Jl. Ganeshano. 10 Bandung, West Java, Indonesia, Ph. +62 22 2502260, Fax +62 22 2534222
saiful@informatika.org

ABSTRACT
In this paper, we describe in-between method found during evolution of minutia-based fingerprint identification from using sequential access method with time complexity $O(n)$ to direct access method with ideal time complexity $O(1)$. We called this method as efficient sequential access method of fingerprint identification since it is still in time complexity $O(n)$. Previous observed sequential access method of fingerprint identification gives detail time complexity $O(n \cdot i \cdot j)$,where n is fingerprint number in database, i is minutia number of input fingerprint, and $n$is minutia number of a fingerprint from database to be matched with. Our efficient sequential access method, without change accuracy performance from previous observation, gives time complexity $O(n \cdot k)$ where $k<i \cdot j$. This k is number of examined minutia-pairs, in index-like minutia-pairs table, that was used to produce best matching score during matching one-to-one. Experiment confirms efficiency of this method compared to the previous sequential access method, through one of the results that gives reduction about 84%, 88%, 87%, 83%, 86%, 85%, 84%, 86%, and 88% on average processed minutia-pairs with impact on higher speed about 5%, 17%, 16%, 8%, 15%, 11%, 13%, 14%, and 18%on average matching time at tested database FVC 2000 DB1A 100, FVC 2000 DB1A 400, FVC 2000 DB1A 800, FVC 2002 DB1A 100, FVC 2002 DB1A 400, FVC 2002 DB1A 800, FVC 2004 DB1A 100, FVC 2004 DB1A 400, and FVC 2004 DB1A 800, respectively.

Keywords
Accuracy, fingerprint, identification, matching, searching, speed.

1. INTRODUCTION
A fingerprint identification system recognizes an individual by searching the entire enrolment templates in database for a match. It conducts one-to-many comparison/matching to establish if the individual is present in the database and if so, returns the identifier of the enrolment reference that matched. In an identification system, the system establishes a subject's identity (or determines that the subject is not enrolled in the system database) without the subject having to claim an identity [1].

Figure 1 shows daily activities of a fingerprint identification system that consists of two main stages, i.e. enrolment and matching. On large database (could be hundreds millions of fingerprints), identification demands real time result so it is important to optimize its speed with respect to its accuracy by tuning up all of those main stages.

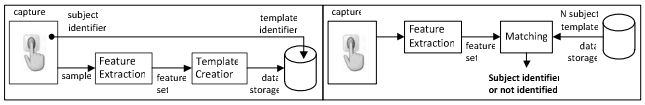


Figure 1. Fingerprint identification system: a) enrolment; b) matching

The identification/searching process can be speeded up by reducing the number of comparisons that are required to be performed. In a general case, intrinsic information of the fingerprint samples has to be used for an efficient retrieval. Searching involves lookup or indexing operations (finding the *value* associated with a *key*) using index. Prudent use of index can make searching faster by eliminating the need to sort (almost always the ultimate goal of sorting is to organize a search) and thus reducing I/O cost [2]. Searching involves sequential or direct access of data. Sequential access is the concept of accessing (or reading) records from a table in sequential order, i.e. from the top to bottom, one after another. Direct access is the concept of accessing specific records from a table in no particular order by specifying which row(s) to be read. The row(s) value possibly comes from processing of specific data representation, like two-dimension quad-tree, kd-tree, and range-tree [3].

Although hypothetically, direct access of fingerprint identification method would outperforms its sequential access method, it is however still important to know some aspects of sequential access method as a base for improvement. Sequential access method of fingerprint identification is the most natural process of fingerprint identification that basically conducts one-to-many matching between input fingerprint (probe) and enrolled fingerprints (candidates) in the database. Because of method's simplicity, method's most natural process, and method's strong foundation for improvement, a disquisition of this method was taking place based on [4] [5] and using [6] as a basic framework. While [4] gives translation- and rotation- variant features (i.e. minutia's absolute coordinate and absolute orientation), [5] gives translation-invariant and rotation-variant features in local-star topology (i.e. reference-minutia – neighbor-minutia edge length, and reference-minutia's absolute orientation), [6] gives translation- and rotation- invariant features in local-star topology (i.e. reference-minutia– neighbor-minutia edge length, reference-

minutia's relative orientation, and neighbor-minutia's relative orientation), this paper observed efficient methodbased on [6] to improvespeed without reducing identification accuracy.

Our efficient sequential access method of fingerprint identification uses fingerprint's local minutiae-based features without pre-selection stage (classification process to produce pre-defined classes). Research design on this method was described on Section 2 with several subsections, i.e.: data structure, sequential access, the algorithm, the improved algorithm as concern of this paper, features computation and comparison, and similarity score.Section 3 gives result and analysis. Section 4 concludes with several points.

2. RESEARCH DESIGN

2.1 Data Structure

Fingerprint identification uses Abstract Data Type (ADT) called *Person* which is primarily a way to group multiple fingerprints belonging to one person. *Person* consists of object *Fingerprint* that contains basic information about the fingerprint, i.e. *Image* that is used to perform template extraction, and *Template* that is used for identification. *Image* is in raw image format that must be set before generating valid *Template*. The format of this image is a simple raw 2D array of bytes. Every byte represents shade of gray from black (0) to white (255). *Template* is an abstract model of the fingerprint that is serialized in a compact binary format. Once the *Template* is generated, *Image* can be set to null to save space. Figure 2 shows approximation of memory space requirement for template contains fingerprint intrinsic information. Templates are better than fingerprint images, because they require less space and they are easier to match than images.

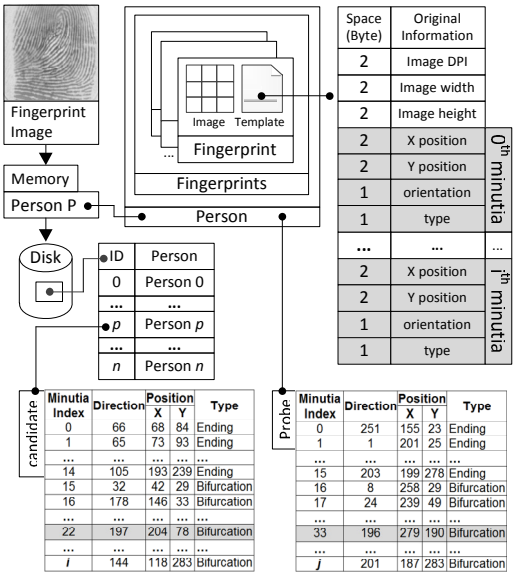


Figure 2. Data structure for fingerprint identification

Object *Person* is designed to be easy to serialize in order to be stored in binary-format (BLOB) under *Person* attribute in database. This binary-format attribute is indexed with numeric-format ID attribute. So here, we need to determine what *key*-like information can be used to construct data structure for searching. Additional requirement, they should be translation- and rotation-invariant for reliability related to identification accuracy. Data structure at Figure 2 do not contains translation- and rotation-invariant information (consider as primitive feature), but that information can be used to generate translation- and rotation-invariant *key*-like information (consider as derivative feature).

2.2 Sequential Access

Figure 3a shows basic approach of sequential access method of fingerprint identification [7]. The matching is a kind of associative array with similarity measure. Probe fingerprint template on input is compared sequentially with database's candidate fingerprint templates stored in the associative array. Candidates are sorted by result of the similarity measure, and the closest match is returned.

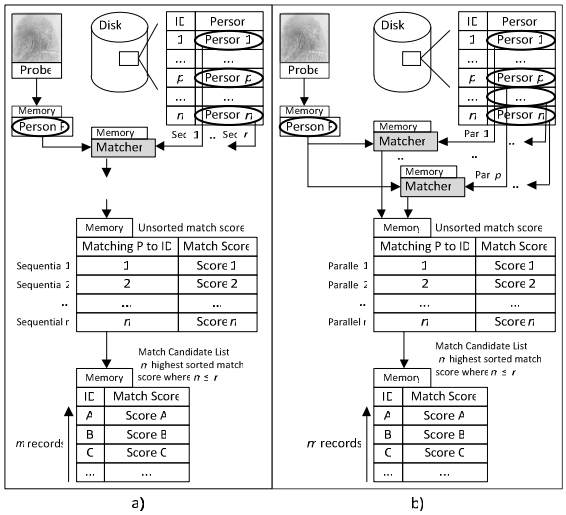


Figure 3. Sequential access method of fingerprint identification: a) basic approach; b) enhancement approach

Enhancement of basic approach (Figure 3b) utilizes *p*cores of processor so there will be *p* concurrent tasks (parallel process) to produce similarity score between probe fingerprint template and different candidate fingerprint template in associative array.

2.3 The Algorithm

Figure 4 shows algorithm used bysequential access method at Figure 3.This algorithm works integrated with ADT called Matcher. Matcher consists of several objects, i.e. ProbeIndex and EdgeTable to store derivative features belong to probe and candidate fingerprint, respectively. The others objects, i.e. MinutiaPairing, EdgeLookup, and PairSelector work together during edge pairing process to produce the pair of the longest paths belong to probe and candidate fingerprint. Pairing of the longest paths will give the best similarity score as indicator for the best similarity between candidate and probe fingerprint. During pairing, some parameters value saved by MinutiaPairing, similar edges belong to probe and candidate saved by EdgeLookup, and decision for the next minutia pair to be travelled handled by PairSelector. Several points about the algorithm:

1. The gray-colored main algorithm comprised of several sub-process, as shown by Figure 4.
2. Small size translation- and rotation- variant minutia information (primitive features) are persisted in database, i.e. its location (*x*,*y*) and orientation *t*, as shown by Figure 2.

3. Relative big size translation- and rotation- invariant minutia information (derivative features) from primitive features are compute in dynamic run-time, i.e. its length of minutiae-edge d, and relative angle between minutiae – minutiae-edge (β_1 and β_2) as shown later by Figure 6b.
4. Derivative features are constructed at object EdgeTable belong to object Matcher. One for the probe fingerprint (encapsulated by ProbeIndex) and one for candidate fingerprint to be matched against.
5. There is global computation time $O(n)$ for identification where n is number of candidate fingerprint templates in database. For each matching between a probe and candidate fingerprint, there is local computation time $O(i \cdot j)$ where i is number of probe's minutia index and j is number of candidate's minutia index, as shown by Figure 2. Pair of i and j is known as pair of minutia index (minutia-pair).
6. The output will be pair of the longest paths, consists of probe path and its counterpart candidate path (shown laterby Figure 7a). This pair of longest paths has its own root that consists of probe's and candidate's minutia index, as shown by gray record at Figure 2, as final objective of fingerprint matching. Pairing of the longest candidate path, start from its root, accumulates a similarity score.

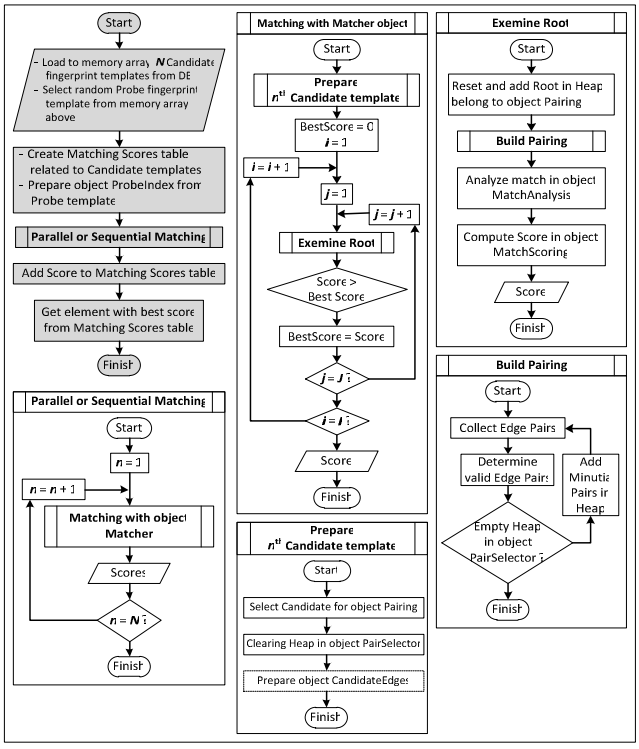


Figure 4. Sequential access algorithm

2.4 The Improved Algorithm

The improved algorithm built our efficient sequential access method of fingerprint identification as proposed by this paper.Several points about the improved algorithm (Figure 5):

1. Improvement was comprised of modules with broken line. Shadow boxes represent modules without any change. There is also module movement shown by bended gray arrow.
2. There is still global computation time $O(n)$ for identification where n is number of candidate fingerprint templates in database. For each matching between a probe and candidate fingerprint, there is improvement in local computation time $O(k)$ where k is length of minutia-pairindex-like table.
3. This candidate's minutia-pairs index contains set S of selected $k$minutia-pairs $\mathbf{p}_s$ which was subset of set A of all $i \cdot j$ $\mathbf{p}_a$ minutia-pairs: $S = \{ \mathbf{p}_s | k \le i \cdot j \}$.
4. Selected $k$minutia-pairs $\mathbf{p}_s$, to put into Index, was chosen from computation of set A which has at least one similarity of its reference-minutia – neighbor-minutia edge length, reference-minutia's relative orientation, and neighbor-minutia's relative orientation.

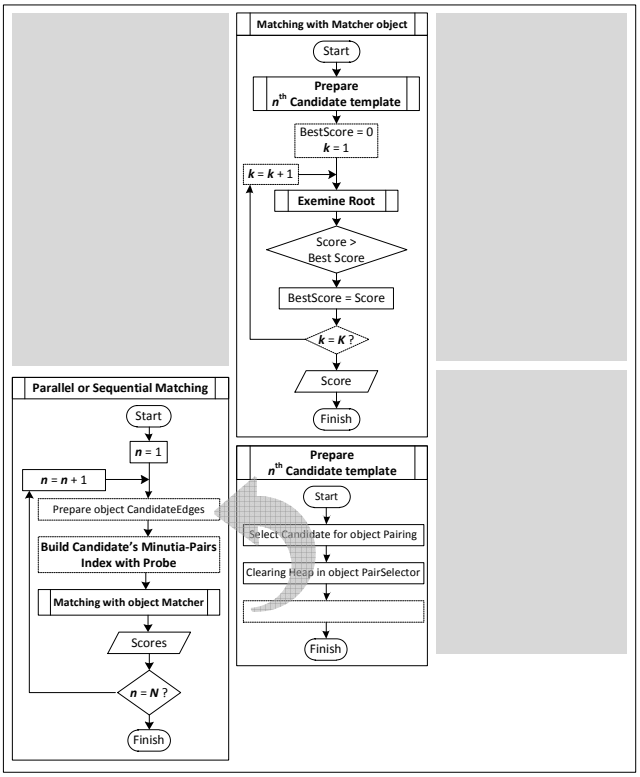


Figure 5. Efficient sequential access algorithm

2.5 Features Computationand Comparison

2.5.1 Computation

Computation of minutia-based local features based on the concept of local structure by using graph notation. The local-star topology(Figure 6a) associated with the minutia $\mathbf{m}_a$ for a given distance d_{max} is the graph $S_a = (V_a, E_a)$ consisting of:

1. The set of vertices V_a containing all minutiae $\mathbf{m}_b$whose spatial distance d from $\mathbf{m}_a$is less than or equal to d_{max}: $V_a = \{\mathbf{m}_a | sd(\mathbf{m}_a, \mathbf{m}_b) \le d_{max}\}$.
2. The set of edges $E_a = \{\mathbf{e}_{ab}\}$, where $\mathbf{e}_{ab}$is the edge connecting minutia $\mathbf{m}_a$with minutia $\mathbf{m}_b$in V_a.$\mathbf{e}_{ab}$is labeled with a 5-tuple $(a, b, d(\mathbf{m}_a, \mathbf{m}_b), \beta_a, \beta_b)$, where $\beta_a$is the reference-minutia's relative orientationand $\beta_b$is the neighbor-minutia's relative orientation.Relative orientation is computed as difference between minutia's absolute orientation (primitive feature) and edge's absolute orientation (derivative feature).

Figure 6b illustrates computation of translation- and rotation-invariant minutia information (derivative features) from primitive features. Two minutiae connected by line construct edge. From left edge of Figure 6b, first minutia *a* is in the upper right and is depicted by the dot representing location (x_a, y_a) and the arrowed line pointing down representing absolute orientation t_a. A second minutia *b* is in the lower left with absolute orientation t_b pointing down and to the left.

To account for relative translational position, the distance d_{ab} and relative angles (between each minutia's orientation and the intervening line between both minutiae) are computed between the two minutia locations. This derivative feature will remain relatively constant between corresponding points on two different finger impressions regardless of how much translation and rotating may exist. In Figure 6b, the angle θ_{ab} of the intervening line between minutia *a* and*b* is computed by taking the arctangent of the slope of the intervening line. Angles β_a and β_b are computed relative to the intervening line by incorporating θ_{ab} and each minutia's absolute orientation *t*.

For each minutia pair comparison, an entry is made into an edge table (Figure 6c), i.e. consists of *key*-like information {d_{ab}, β_a, β_b}. Entries are stored in object EdgeTable which is encapsulated by object Matcher. In ascending order of distance, the table is trimmed at the point in which thresholds (a maximum distance and a maximum neighbor) are reached. Making these measurements among minutiapairs, an edge table must be constructed for pair of probe fingerprint and candidate fingerprint we want to match with.

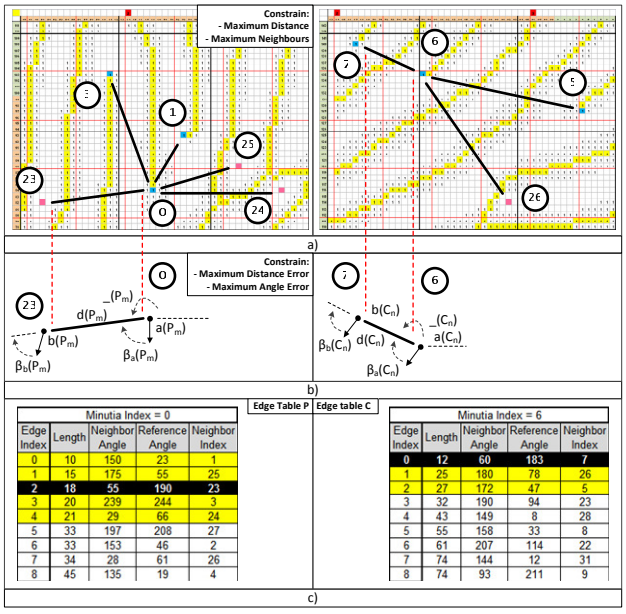


Figure 6. a) Portion of two slight different impressions of the same fingerprint. Left: probe, right: candidate; b) Derived features of minutiae points; c) Sample of fingerprint's *key*-like information on edge table, where yellow records related to displayed entries, and black records indicate similar entries

2.5.2 Comparison

Derivative feature comparison takes the minutia edge tables from two separate fingerprints and look for similar entries between two tables. Figure 6a shows portion of two impressions of the same fingerprint, with slight differences in both translation and rotation.

The left print represents a probe impression in which all of its minutiae have been pair-wised computed with relative measurements stored in edge table P (Figure 6cshows the sample). The relative measurements computed from the particular pair of minutia (Figure 6b) have been stored as the m^{th} entry in table P, denoted P_m. The index of the upper right minutia is stored in table entry $P_m$and is referenced as $a(P_m)$, while the distance between the two minutiae is also stored in table entry $P_m$and is referenced as $d(P_m)$. The right print represents a candidate impression from database, and uses similar notation, except that all its pair-wise minutia comparisons have been stored in table C, and the measurements made on the two corresponding minutiae in the candidate print have been stored in table entry C_n.

Black records at Figure 6c indicate similar entries. Three tests are conducted to determine if table entries $P_m$and $C_n$are similar. The first test checks to see if the corresponding distances are within a specified tolerance: $\Delta d(d(P_m), d(C_n)) \leq T_d$.The last two tests check to see if the relative minutia angles are within a specified tolerance: $\Delta\beta(\beta_a(P_m), \beta_a(C_n)) < T_\beta$, and $\Delta\beta(\beta_b(P_m), \beta_b(C_n)) < T_\beta$.

2.6 Similarity Score

Similarity score between probe and one candidate from database is obtained through pairing (traversal of the pair)of the longest paths, where each path belongs to the probe and candidate to be match with. Figure 7 shows part of process to calculate similarity score of candidate compared to the probe.

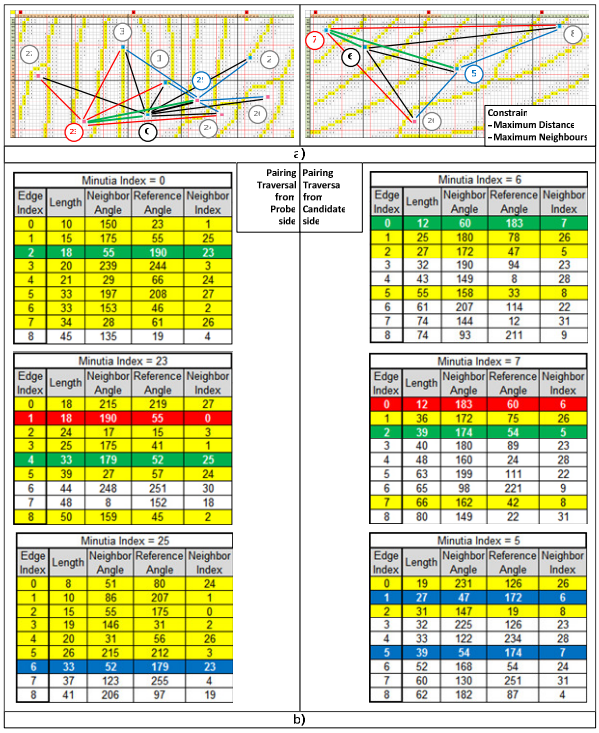


Figure 7. a) The longer the pairing traversal path, the higher the similarity score; b) edge table information

Maximum similarity score produced by accumulating similarity score parameters during pairing of the longest path (maybe

discontinuous path) of both probe and candidate. The pair of green path at Figure 7a actually is not the pair of the longest path but it illustrates the process to find it. These longest paths have its own root (consists of probe's minutia index and candidate's minutia index) that gives the best similarity score as final objective of fingerprint matching.

Figure 7b shows sample of pairing information start from minutia pair [0,6] until minutia pair [25,5]. Yellow records show displayed edge at Figure 7a. Green, red, and blue records are similar edge pair (edge from probe and candidate, respectively) belongs to its minutia pair during pairing.Green records are also special entries that construct pair of the longest path. Several points during processing of similarity score, i.e.:

1. Finding pair of the longest path through comparison with others pair of the longest path during iteration on probe and one candidate fingerprint.
2. After finding pairing of the longest path, its best similarity score was put as a value in unsorted associative array with *Person* ID as a key (Figure 3). *Person* ID represents index (location) of a candidate fingerprint in database. The highest score in associative array become the strong similar candidate among others candidates fingerprint in database.

Some parameters for traversal score computation during pairing of the longest path are minutia-pair count, correct type minutia-pair, and supported minutiae pair count. The last two parameters are used for strict filtering of minutiae pair count for better deal with false matches that cause problems with low quality fingerprints.

3. RESULTS AND ANALYSIS

Monte Carlo method was conducted on experiment that relies on repeated random sampling to compute the results. The analysis is most suited to calculation by a computer and tends to be used when it is infeasible to compute an exact result with a deterministic algorithm [8]. The method follows these steps: 1) Define a domain of possible inputs; 2) Generate inputs randomly from a probability distribution over the domain; 3) Perform a deterministic computation on the inputs; 4) Aggregate the results.

In step 1, wedefine the domain of inputs is Set A of Database 1 (DB1A, fingerprint image collected by using small-size and low-cost optical sensors) of Fingerprint Verification Contest (FVC) 2000 [9], 2002 [10], and 2004 [11], each with 300x300, 388x374, and 640x480 pixel size. Then each database was set up with population number of 100, 400, and 800 data (Every Set A of FVC Database has maximum value of 800 data). Database with population number of 100 is selected by omitting all but first impression of 100 volunteers' fingerprints, respectively (each volunteers' left- and right- fore fingers and middle fingers were acquired twice on two different-day session). Database with population number of 400 is selected by omitting second session's four impressions of 100 volunteers' fingerprints, respectively. So here, we have domain of inputs come from nine different databases (differ from FVC year and population number).

Table 1. Computation mode for random input

Mode	Index[1]	Parallel[2]	Note
00	No	No	[1]Minutia-pairs index [2]Multi thread by multi cores [3]Parallel process with reused object Matcher [4]Parallel process without reused object Matcher
01	No	Yes[3]	
02	No	Yes[4]	
10	Yes	No	
11	Yes	Yes[3]	
12	Yes	Yes[4]	

In step 2, we generate arbitrary number (a hundred) of random inputs per database, and followed by step 3 where we perform different computation mode, as shown by Table 1, on each random input for searching/matching itself on database.Finally in step 4, we aggregate the results per database, i.e., the approximation of some observed parameters with their average value (*Avg*), standard deviation value (σ), minimum value (*Min*), and maximum value (*Max*), respectively. Standard deviation value was calculated using formula with Bessel's correction [12] below (the use of *N*-1 instead of *N*):

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

where $\{x_1, x_2, \ldots$ are the observed parameter of the random sample items and $\bar{x}$ is its mean value (*Avg*). To get an accurate approximation for all those values, this procedure has two common properties of Monte Carlo method. First, the inputs should truly be random. Second, there should be a large number of *N* random inputs (step 2). We conducted this experiment on Microsoft® Windows™ 7 machine with RAM 4 GB and Intel® Core™ i5-2430M (Quad Cores) @ 2.20 GHz.

Figure 8 shows average matching time,as one of observed parameters, on different computation mode. It clearly shows that our efficient sequential access method was outperformed its counterpart computation mode consistently amongnine tested databases (mode 10 outperforms mode 00, mode 11 outperforms mode 01, and mode 12 outperforms mode 02). On the using of parallel processfor the same index mechanism, computation mode *1 always outperform computation mode *2 because of the reuse mechanism of object Matcher (see Figure 4) that gives less computation time rather than creating the new one. And finally, on the comparison of parallel process (multi cores) to sequential process (single core)for the same index mechanism, computation mode *1 or computation mode *2 always outperform computation mode *0 because of concurrent works nature of multi cores.

Figure 9 shows another observed parameter, i.e. average processed minutia pairs on computation modes 01and11 which have highest speed among others computation modes during matching random input for overall nine tested database.

In general, because of the reduction of average processed minutia pairs (above 80% as shown by Table 2), our efficient sequential access method of fingerprint identification gives higher average matching time (up to 18%), if compared to normal sequential access method of fingerprint identification.

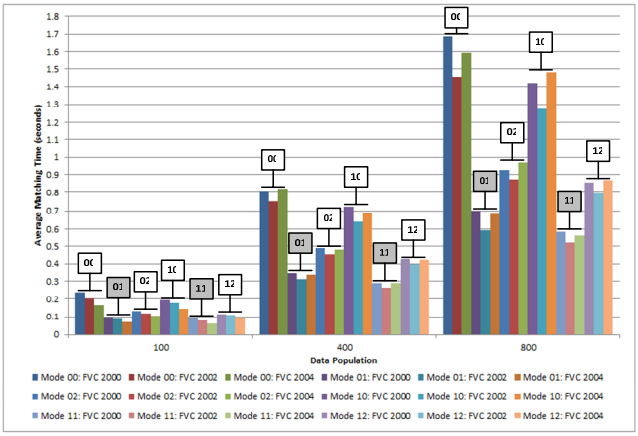


Figure 8. Average matching time on different computation mode for overall nine tested databases

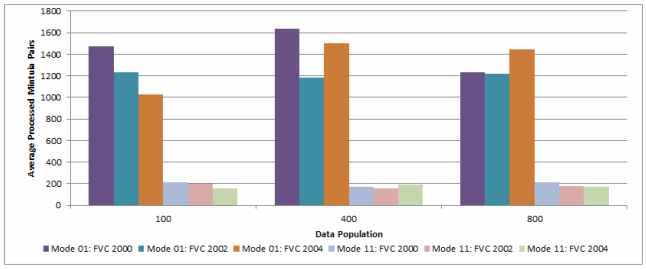


Figure 9. Average processed minutia pairs on computation mode 01 and 11 for overall nine tested databases

Table 2. Observed parameters efficiency

	Mode = 01		Mode = 11		Efficiency (%)	
Db[1]	AMT[2]	APMP[3]	AMT	APMP	AMT	APMP
1	0.098	1411	0.092	212	5.37	84.984
2	0.351	1474	0.290	172	17.24	88.340
3	0.700	1636	0.585	212	16.45	87.032
4	0.086	1230	0.079	198	8.198	83.932
5	0.310	1183	0.261	161	15.71	86.428
6	0.593	1216	0.522	182	11.94	85.029
7	0.071	1030	0.062	157	13.05	84.709
8	0.338	1500	0.288	195	14.74	86.985
9	0.685	1449	0.558	168	18.58	88.403

[1]Db: 1 = FVC 2000 DB1A 100, 2 = FVC 2000 DB1A 400, 3 = FVC 2000 DB1A 800, 4 = FVC 2002 DB1A 100, 5 = FVC 2002 DB1A 400, 6 = FVC 2002 DB1A 800, 7 = FVC 2004 DB1A 100, 8 = FVC 2004 DB1A 400, 9 = FVC 2004 DB1A 800.
[2]AMT = Average Matching Time in seconds.
[3]APMP = Average Processed Minutia Pairs.

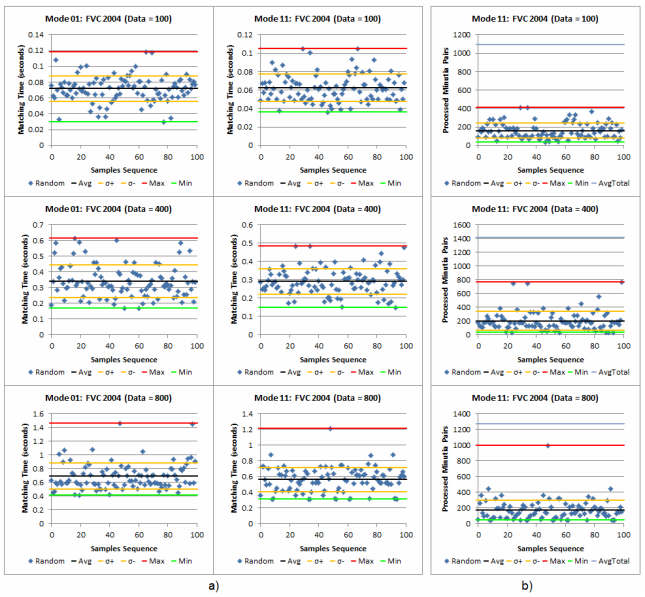


Figure 10. a) Matching time fromcomputation mode 01 (left) and 11 (right) of FVC 2004DB1A; b) Processed minutia pairs for computation mode 11 of FVC 2004 DB1A

Figure 10 shows in detail of two observed parameters, i.e. matching time and processed minutia pairs, per sample for arbitrary selected FVC 2004 DB1A. It shows sample's matching time from computation mode 01 and 11 (Figure 10a) and sample's processed minutia pairs from computation mode 11 (Figure 10b) with their average value (*Avg*), standard deviation value (σ), minimum value (*Min*), and maximum value (*Max*), respectively. Figure 10b also shows Average total minutia pairs (*AvgTotal*) that will be used by normal sequential access method of fingerprint identification. Previous results for overall nine tested database, were computed through this mechanism.

4. CONCLUSION AND FUTURE WORK

We have observed efficient sequential access method of fingerprint identification as in-between method during evolution of fingerprint identification from using sequential access method to direct access method. The experiment confirms its efficiency through observed average matching time and processed minutia-pairsparameters. Future work on direct access method, need to be observed for time complexity $O(1)$ rather than $O(n)$.

5. REFERENCES

[1] Maltoni, D., Maio, D., Jain AK, and Prabhakar, S. 2009. Handbook of Fingerprint Recognition. Second Edition.Springer -Verlag,London.

[2] Snell, G.P. 2006. Think FAST! Use Memory Tables for Faster Merging. In SAS Users Group International 31 Proceedings(San Francisco, USA, 2006). SUGI'06.

[3] Samet, H. 1990.The Design and Analysis of Spatial Data Structures. Addison-Wesley, New York, NY.

[4] Watson, C.I., Garris, M.D., Tabassi, E., Wilson, C.L., McCabe, R.M., Janet, S., and Ko, K. 2001.User's Guide to Export Controlled Distribution of NIST Biometric Image Software. National Institute of Standards and Technology.

[5] Ratha, N.K., Pandit, V.D., Bolle,and R.M., Vaish, V. 2000. Robust Fingerprint Authentication Using Local Structural Similarity. Proc. Workshop on Applications of Computer Vision, 2000, 29–34.

[6] Vazan, R. 2009. SourceAFIS - Open Source AFIS on SourceForge. http://sourceforge.net/projects/sourceafis/

[7] G. Indrawan, B. Sitohang, S. Akbar, "Review of Sequential Access Methodfor Fingerprint Identification", TELKOMNIKA Indonesian Journal of Electrical Engineering, 2012, 10(2), in press.

[8] H.L. Anderson,"Metropolis, Monte Carlo and the MANIAC", Los Alamos Science, 1986, 14, 96–108.

[9] D. Maio, D. Maltoni, R. Cappelli, J.L.Wayman, A.K. Jain,"FVC2000: Fingerprint Verification Competition", IEEE Transactions on Pattern Analysis Machine Intelligence, 2002, 24(3), 402-412.

[10] Maio, D., et all. 2002. FVC2002: Second Fingerprint Verification Competition. Proc. 16th Int'l Conf. Pattern Recognition (2002). 811-814.

[11] Maio, D., et all. 2004. FVC2004: Third Fingerprint Verification Competition. Proc. Int'l Conf. Biometric Authentication (2004). 1-7.

[12] Montgomery, D.C., and Runger, G.C. 2003. Applied Statistics and Probability for Engineers. Third Edition. Wiley and Sons, New Jersey, NJ.

Randomized Heuristics Algorithm for Container Loading Problem: A Case Study

Djoni Haryadi Setiabudi
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
+62318439040
djonihs@petra.ac.id

Gregorius Satia Budhi
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
+62318439040
greg@petra.ac.id

Alex Chandra Suryana
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia

ABSTRACT
At a cargo company the problem is how the process of loading can be made efficiently in order to include as many items as possible. To make the cargo loading process more swiftly, our research is to create an application to simulate the process of cargo loading. This simulation is performed on one of the cargo company, namely PT SLI as a case study. This company need a cargo loading simulation application which is at the end can help the company in the day to day cargo loading process.

The Process from this application starts from entering the data of container and goods. The next step is the process of simulating cargo loading using Randomized Heuristics algorithm, which will produce a picture layout of goods clearly. This application is made by using 3 kinds of software, there are Microsoft Visual C# 2010, Microsoft Visual C++ and Microsoft SQL Server 2008 as data storage.

From the test results of the application, can be concluded that the application can determine the number and the type of container needed in the cargo loading process. The number of container used can be minimized so that it can reduce the shipping cost and also be able to solve the multi-drop destination.

Keywords
Cargo loading, randomized heuristics algorithm, Container Loading Problem (CLP)

1. INTRODUCTION
The general container loading problem in the cargo company is to put as many as three-dimensional rectangular boxes into a rectangular container of fixed dimensions and minimized the empty spaces among boxes. Each box must be completely contained in the container and cannot overlap with the other boxes. The problem may arise is that the size of the boxes are not exactly rectangular, so that the real spaces should be reserved are more than the size of the boxes. The other problems that occur in the cargo company is the difficulty of determining the number of truck / container that is used in the delivery of goods. Often trucks are used more than that should be required so that it increase the shipping cost to be paid by the company. Unlike the container loading problem in global logistics that the majority of goods for a container are of the same size and for one client, a truck container used in city logistics often needs to transport goods in different sizes and have more than one destination in one route. Hence, the inner-city truck loading problem considers not only the traditional container loading constraints, but also the loading order of goods for different destinations (like retail outlets and supermarkets), which requires that items can unload without rearranging other items at a destination. This requirement is obvious, since rearranging goods at each destination (or multi-drop point) may increase cost time and labor cost, and may result in goods damage.

The most common heuristic approaches can be classified as wall-building algorithms (Bischoff and Marriott[2], Hemminki[6]), stack-building algorithms (Gilmore and Gomory[5]), guillotine-cutting algorithms (Morabito and Arenales[8]), and cuboid-arrangement algorithms (Bortfeldt and Gehring[3]). Morabito and Arenales introduced a procedure for intelligent graph searches with and/or graphs, while Gehring and Bortfeldt developed a genetic algorithm for container loading problems.

In this paper we developed an application based on heuristic procedure developed by Juraits et.al. [7] and Inner City Multi-Drop by Pan et.al. [9]. We made an application and the simulation process using the data supplied by PT. SLI, a cargo company based on Surabaya, Indonesia.

2. HEURISTIC ALGORITHMS
The wall-building approach, introduced by George and Robinson [4], fills the container in a number of layers across the depth of the container. See Figure 1. This heuristic is based on the normalized layer depths: only layer depths d, which equal some box di-mensions were considered. The depth of a layer must be carefully selected to obtain a good performance. Then the wall is packed in a greedy way as a number of horizontal strips. Every strip is packed by consecutively inserting boxes with the largest ranking. The ranking is based on the smallest dimension of a box, then on the number of remaining boxes of a given type and finally on the length of the largest dimension. [7]

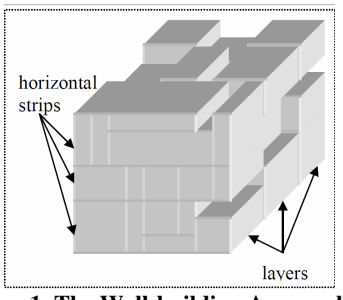


Figure 1. The Wall-building Approach [4]

The major disadvantage of all wall-building heuristic is that they are greedy: making locally optimal choices, however, may lead to

bad final solutions. To obtain a better performance, it is necessary to have some kind of backtracking possibility.

The algorithm presented by Baltacioglu[1] builds walls or layers along any of the six faces of the given container. This heuristic employs both layer packing and wall building approaches. One of the most important methods, used in this algorithm, is a layer-in-layer packing approach: packing a sub-layer into any of the available unused spaces in the last packed layer. See Figure 2.

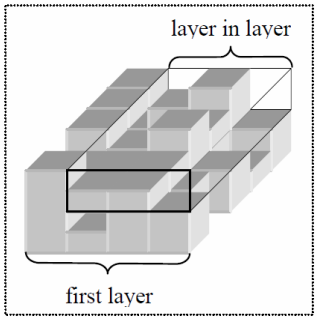


Figure 2. Layer in Layer Packing [1]

3. RANDOMIZED HEURISTIC

The main core of the algorithm is divided the container into several layers. Container and packing process in the layer resolved by randomized heuristics [7].

The working steps of the is randomized heuristics algorithm are :

1. Select the next box with the largest volume available :
   $f_1 \rightarrow \max\{v_i\}$
2. Choose the box with the largest ratio box_ volume/bounding_sphere_volume.
   $f_2 \rightarrow \max\left\{\frac{v_i}{V_i}\right\}$
   Where V_i is the bounding sphere volume.
3. From the rest of every box, is chosen the smallest of width, height and depth, then choose the largest.
   $f_3 \rightarrow \max\{\min\{w_i, h_i, d_i\}\}$
4. Then the container divided into several layers, the layer that has a maximum value in the ranking is chosen for loading the goods to determine the ranking functions layer. The ranking functions are :

4.1 $f_k^1 = \sum_{i=1}^{n} 1\,(w_i = k \cup h_i = k \cup d_i = k)\ for\ k = \alpha, \ldots, \beta$

4.2 $f_k^2 = \sum_{i=1}^{n} 1\,(max\{w_i, h_i, d_i\} = k)\ for\ k = \alpha, \ldots, \beta$

4.3 $f_k^3 = \sum_{i=1}^{n} 1\,(min\{w_i, h_i, d_i\} = k)\ for\ k = \alpha, \ldots, \beta$

5. The fourth function of frequency capture the smallest value of the width, height, and depth, after it is selected that has a maximum value of the remaining box, similar in function to the number (3):

   $f^4 = \max\{\min\{w_i, h_i, d_i\}\}, i = 1 \ldots n$

   It's possible to obtain better results using a mixture of these heuristics. Of each layer has the best value obtained by using a heuristic function and then chose the solutions that deliver better value content. Solution gives better results obtained by the function (4.1) and (5).
6. First of all, every layer is filled using all three mentioned greedy approaches (1-3). After that, one additional heuristic for selecting next box for packing was implemented. A box having at least one dimension equal to the current layer depth is the most valuable; if there is no such box, then the most valuable box is assumed by previous heuristics. Heuristic function (1), (2), (3), and (6) is used to determine which box to be loaded next. To place the laying boxes in the container, it must be a list of places still available. Firstly, the list filled with a point which is the angle of the container (layer). Then we search the box can be placed at this point and enough to put in a container (layer) and should not exceed the layer dimensions. If the box satisfied these criteria, we eliminate the point and add three new points. If the box is not suitable with the laying of a new point, the box is rotated and redo the laying process. If after the box is rotated, we can not found an appropriate point then he box is ignored.

4. SYSTEM DESIGN

Based on the analysis of the problems, the company requires a Cargo Loading application that will facilitate the delivery of goods by the company. This application aims to simulate the loading of goods in a container so that the loading process can run faster and companies can find out the number of trucks required for the delivery of goods.

4.1 Design Criteria

This application will store the container master data, master items, and delivery session, whether it includes the session container (container recorded sessions that will be used) as well as goods session (a session that records the items to be sent) and finally create a procedure to simulate the loading of goods and produce output in three dimensions.

The application will also take into account some important things in the process of loading the goods in accordance the requirement of the cargo company, there are:

- The first input contains data about the container that will be used include the type, length, width and height of container. The second input contains data about the goods to be loaded into the container, consisting of the type, length, width, height and weight of the item.
- The process of data processing using randomized heuristics algorithms.
  Calculate the volume of container that can be used. Rearrange the position of the boxes in the container according the volume of the each box.
- Calculate the remainder volume of container that can be used after loading some boxes.
- If the container still has some empty space, it can continue to entering the box into the container.
- Arrangement of the boxes should be adjusted so that the boxes are not damaged.
- There should be any tolerance value for each item because of the size of the boxes is not precision.

- The maximum number of stack items should not be exceeded. There should be any arrangement of the boxes so that the right and the left of the container are balanced.
- The output should produce a picture of the layout of a 3 dimensional items.

4.2 Design Flowchart

Flowchart of the system design can be seen in Figure 3. The first input is the data input for *Master Container*, a container type and dimensions, including length, width and height of the container and the latter is the maximum weight of the container. The second input is the *master data of goods*, in the form of the name, length, width, height, weight, tolerance, maximum stack and maximum weight of goods. The third input is the input for *master session*, the session makes notes on the delivery of goods, namely the date of shipment and delivery details which consists of the session name and destination of the session delivery. The next input is an input data for *Good session* which contains the goods name, priority placement of goods in the container based on shipping destination and the number of items to be loaded. The last input is the input for *session Container* consist of types of goods, priorities for the container and the number of container available to be used.

After the process of input the data, the next thing to be done is to process the data generated from C # to C + + which is the process of removing data that will be used to simulate the process of loading items into a text file. The next process to be performed is Simulate Cargo Loading Using Randomized Heuristic Algorithm, in this process the text file generated is read by the C++, then the simulations using randomized heuristics algorithms is to be done. Then go to the last process that generates a 3D view with C + +, where the layout of items in a container after the simulation is described on the screen in three dimensions.

5. EXPERIMENTAL RESULTS

Testing is done by running the application that has been made, whether it has been able to resolve problems faced by the company every day.

Application testing is done using the original data obtained from the cargo company. Simulations carried out using two types of container that has a different size. The shipment information can be seen in Figure 4. The information includes type of shipment, the sender and destination address. This shipment is multi-drop delivery.

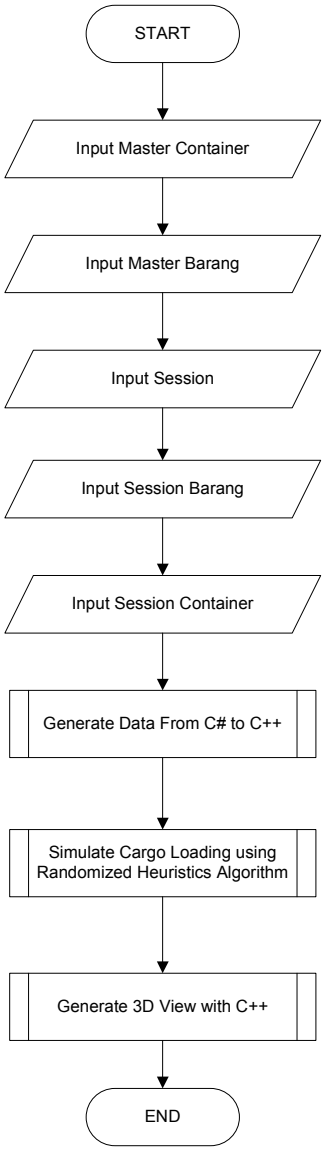


Figure 3. Flowchart of Cargo Loading Application

The First Trial

Tests conducted using the real data of goods and containers from the company, which will then be followed by the simulation process of loading goods into the container. The original data can be seen in Figure 5. Because it is a multi-drop delivery so the priority of goods is adjusted according the destination. Items to be sent to the first destination given the first priority, while the goods to be sent to the second place rated the second priority. Goods with the second priority will be loaded first and placed into a container of, goods with first priority will be loaded at the next turn and placed in front of the second priority items. This sequence is done so that at the time of unloading the goods do not need to unload the second prioriy goods. Figure 6 shows the output result of the first trial of simulation from the original data.

Type of Shipment	: MULTI DROP 2 destination.
Sender	: PT Philip Morris Indonesia
	ID01:Bekasi Manufacturing, Bekasi Manufacturing
	Kawasan Industri MM2100 Bekasi
Receiver (1)	: PT HANJAYA MANDALA SAMPOERNA TBK.
	ZD4A:Depo Jakarta Karawang
	Jln. Permata II Lot BB-3,4B,7 Karawang
Receiver (2)	: PT HANJAYA MANDALA SAMPOERNA TBK.
	ZD4E:ADO Tangerang
	Jl. Sinar Hati Blok N No. 88 Tangerang

Figure 4. The Shipment Information

Container Data:

No.	Container Type	L (cm)	W (cm)	H (cm)	Max Weight (kg)	Priority	Nb
1.	Hino FL 210 Modif Long Wing Box Muntjul Diamond.	950	237	245	13000	1	1

Data of Goods:

No.	Goods Name	L (cm)	W (cm)	H (cm)
1.	MARLBORO GOLD LIGHTS KS BOX 20	25	46	58
2.	MARLBORO BLACK MNT KS BOX 20	25	46	58
3.	MARLBORO (RED UPGRADE) KS BOX 20	25	46	58
4.	MARLBORO MENTHOL LIGHTS MNT KS BOX 20	25	46	58

Data of Goods (Cont):

No.	Wgt (Ons)	Tol	MSN	MSX	MSZ	MSW (Ons)	Priority	Nb
1.	144	1	7	8	9	2600	1	266
2.	144	1	7	8	9	2600	2	32
3.	144	1	7	8	9	2600	2	131
4.	144	1	7	8	9	2600	2	20

Figure 5. Original Data From The Cargo Company

Based on the output of the first trial, it can be concluded that:

- Overall of the goods can be carried out using one container with the size 950 x 237 x 245 cm.
- The remaining weight of the container = 130000-64656 = 65 344 ounces = 6534.4 Kg.
- The remaining volume of container = 25 m3.
- Container can still be loaded with the goods again because it still has some empty space and has not exceed the maximum weight.

The Second Trial

In the second trial of the simulation using the original data, the testing is done by changing the type of container used to load the goods. Container previously used that has the size 950 x 237 x 245 cm will be replaced with a smaller container whose size is 725 x 250 x 247 cm on Figure 7.

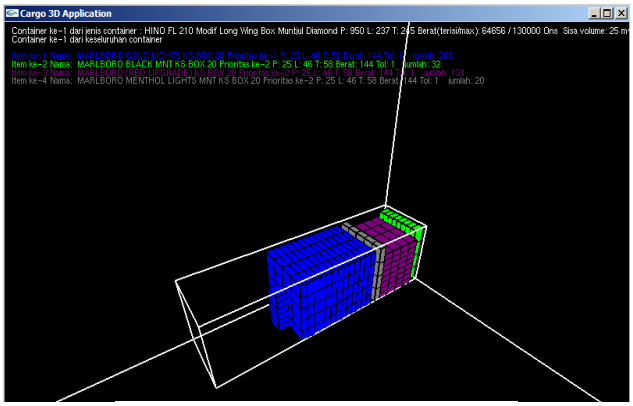


Figure 6. The Output of The First Trial

Container Data – The Second Trial:

No.	Container	L (cm)	W (cm)	H (cm)	Max Weight (kg)	Priority	Nm
1.	HINO RANGER 176 Wing Box Antika Raya	725	250	247	13000	1	1

Figure 7. The Second Container

Figure 8 shows the output result of the second trial of simulation from the original data.

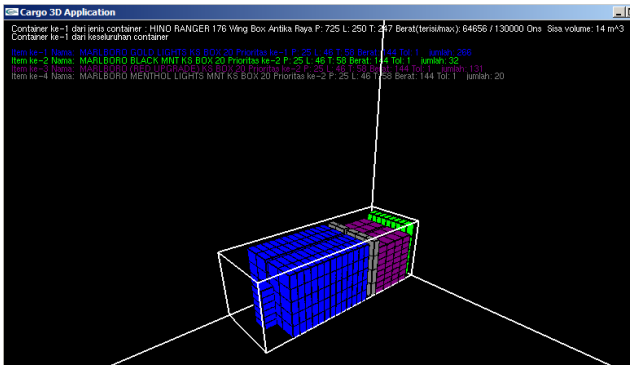


Figure 8. The Output of The Second Trial

Based on the comparison from the output of the first trial and second trial, it can be concluded that:

- Settings for Multi Drop items are correct, the first priority items ahead of second priority items, according to the shipping route.
- The smaller container whose size is 725 x 250 x 247 cm was able to load the entire items. All the goods that were previously loaded using container sized 950 x 237 x 245 cm, once tested it only need the container of 725 x 250 x 247 cm to load all the stuff. It means the company can save the delivery cost than using the large-sized container.
- The remaining weight of smaller container is 6534.4 Kg.
- The remaining volume of smaller container is14 m3.

6. CONCLUSIONS

- The design of the algorithms used to simulate the process of loading the goods have been able to improve efficiency. It can

be proved by testing with the original data that has been done, where the loading of the goods that previously used container size 950 x 237 x 245 cm, after the trial overall goods can be loaded on a smaller container size 725 x 250 x 247 cm. Using a smaller container will decrease the operational cost.

- Based on the results of the testing of multi-drop destination, the process of loading the goods on the simulation of multi-drop delivery to the correct places, where the placement of goods in the container is in conformity with the order of the place of delivery of goods. Items to be sent to the first placed in front of the container near the door, while the goods to be shipped over to the next placed in a container in accordance with the order of the transfer of goods.

7. REFERENCES

[1] Baltacioglu, E. 2001. The distributor's three-dimensional pallet-packing problem: a human intelligence-based heuristic approach. Working paper, *Air Force Institute of Technology*, Wright-Patterson Air Force Base, Ohio, USA

[2] Bishoff, E.E., Marriot, M.D. 1990**.** A comparative evaluation of heuristics for container loading. *European Journal of Operational Research*, 44, 1990, 267-276.

[3] Bortfeldt, A., Gehring, H. 1997**.** Applying Tabu Search to Container Loading Problems. *Operations Research Proceedings*, 533-538.

[4] George, J.A., Robinson, D.F. 1980**.** A heuristic for packing boxes into a container. *Computers and Operations Research*, 7, 1980, 147-156.

[5] Gilmore, P.C., Gomory, R.E. 1965. Multistage cutting stock problems of two and more dimensions. *Operations Research*, 13, 94-120.

[6] Hemminki. 1994**.** Container loading with variable strategies in each layer. ESI-X, *EURO Summer Institute*, Jouy-En-Josas, France.

[7] Juraits M., et.al. 2006. A Randomized Heuristic For The Container Loading Problem: Further Investigations. *Information Technology and Control*, Vol. 35, No.1

[8] Morabito, R., Arenales, M. 1994. An AND/Orgraph approach to the container loading problem, *International Transactions in Operational Research*, 1, 256-267.

[9] Pan L., Chu, S.C.K., Han, G., Huang, J.Z. 2011. A Heuristic Algorithm for The Inner-City Multi-Drop Container Loading Problem. *International Journal of Operations Research and Information Systems*, 2(3),1-19.

Facial Emotional Expressions Synthesis using Radial Basis Function Network

Endang Setyati
Sekolah Tinggi Teknik Surabaya
Informatic Department
Surabaya, Indonesia
+62315027920
endang@stts.edu

Yoyon K. Suprapto
Institut Teknologi Sepuluh Nopember
Electrical Engineering Department
Surabaya, Indonesia
+62315994251-54
yoyonsuprapto@ee.its.ac.id

Mauridhi Hery Purnomo
Institut Teknologi Sepuluh Nopember
Electrical Engineering Department
Surabaya, Indonesia
+62315994251-54
hery@elect-eng.its.ac.id

ABSTRACT
Emotion recognition through the computer-based of facial expression has been an active area of research in the literature for a long time. In this paper we develop a facial emotional expression synthesis (FEES) techniques based on the facial features extracted from facial characteristic points (FCPs) in frontal image sequences. In order to synthesize such facial expression, FCPs can be used as landmarks. These FCPs are feature points that represent significant movements during the generation of expression. We try to simulate people's expressions artificially using Radial Basis Function Network (RBFN. The experimental result from classifier, success rate was about 91.57% using RBFN classifiers.

Keywords
Facial emotional expression synthesis; facial features extracted; facial characteristic points; radial basis function network

1. INTRODUCTION
Face plays an essential role in interpersonal communication. Facial expressions play an important role in non-verbal communications. Many applications for teleconferencing, human computer interface and computer animation require realistic reproduction of facial expressions.

The interest in computational models of emotion and emotional expressions has been steadily growing in the agent research community. Several psychologists have acknowledged the role of emotions in intelligence [1].

In 1971, Ekman and Friesen [2] postulated six primary emotions that each posses a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions. They seem to be universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise, and anger. The method of recognizing the 6 universal facial expressions using neural network is discussed.

Most emotion theorist [3] emphasize the involuntary nature of emotional experience, ignoring those instances when people choose to generate an emotion through reminiscence or by adopting the physical actions associated with a particular emotion (e.g., speaking more softly to deintesify anger or smiling to generate enjoyment).

[4] believed there exist a relationship between facial expression and emotional state. There is a small set of basic emotions that can be expressed distinctively from one another by facial expressions. For instance, when people are angry they frown and when they are happy they smile.

Facial expressions [5] are the result of facial muscle actions which are triggered by the nerve impulses generated by emotions. The muscle actions cause the movement and deformation of facial skin and facial features such as eyes, mouth, and nose. We can use optical flow to estimate facial muscle actions which can then be used in recognizing a facial expression.

[6] apply a 234x50x6 back-propagation neural network for classification of expression into one of six basic emotion categories and their strength. Then they generate the facial position information and it is input into the input units of neural network, networ learning is done by back-propagation algorithm and recognition test is carried out. For six basic facial expressions, the correct recognition ratio is found to about 90%.

Suppose we are given a 2D grayscale expressionless face image of a person, i.e., a face image without any expression of emotion [7], how can we synthesize different expression of that person? One of the ways to synthesize facial expressions is to find the approximate displacement of prominent facial feature points. Moreover, [8] often cannot obtain accurate facial landmark displacement information due to inherently inaccurate input data. This is because: (1) it is hard to generate a set of standardized expressions, e.g., each person may smile differently, (2) it is hard to produce accurately the precise degree of a particular expression, e.g., how to generate a 20% smile?, and (3) it is difficult to mix various facial expressions, e.g., how to gesture a happy and sad face?

[9] in synthesizing facial expressions include texture mapping approach to 3D facial image synthesis and [10] use of 3D model of facial muscles and tissues. An alternative approach has been investigated by [11], which demonstrated the use of RBF in interpolating the anchor points for 2D image warping, which can be applied to synthesize facial expressions. However, it is does not provide a mechanism to determine the appropriate destination of the anchor points for each particular facial expression.

Human face has several unambiguous features: eyebrows, eyes, mouth, nose, and face outline. [13] extract three main features: eyebrows, eyes, and mouth. After extracting these feature, they are able to get 30 points of the FCPs.

[14] proposed a hierarchical model of RBFN to classify and to recognize facial expressions. This approach utilizes Principal Component Analysis as the feature extraction process from static images. This research is to develop a more efficient system to discriminate 7 facial expressions. They achieved the correct

classification rate above 98.4% which is overwhelmingly distinguished compared to other approaches.

[15] develop a facial expression recognition system, based on the facial features extracted from FCPs in frontal image sequences. Selected facial feature points were automatically tracked using a cross-correlation based optical flow, and extracted feature vectors were used to classify expressions, using RBFN and FIS. Success rates were about 91.6% using RBF and 89.1% using FIS classifiers.

In this paper, we proposed two systems for classifying of the facial expressions from The Japanese Female Facial Expression (JAFFE) Database [17]. 7 features extracted from 30 feature points and from a feature vector for each expression. These feature vectors were used to training a RBFN classifier to classify input feature vectors into one of the six basic emotions.

2. INFORMATION OF FACIAL EXPRESSION

2.1 Facial Characteristic Points

FCPs carry the information about the position and shape of these three features. According to the study of the Ekman and Friesen [12], almost all facial expressions of human face are described by combination of 46 basic movements of facial muscles and these basic movements are called Action Units (AUs). 30 AUs are directly associated with movement of eyes, eyebrows, and mouth. That is why the information expressing movement of eyes, eyebrows, and mouth is desirable for machine recognition of facial expressions. The information of each of 6 basic facial expressions is obtained by subtracting the FCPs coordinates of normal facial expression from those of facial expressions.

In [6] are confined in these three components and then determine FCPs which are representative of the boundary between these components and skin. A set of 30 facial landmarks located near the eyes, eyebrows, and the mouth are defined as the FCPs. These points are shown in Figure 1.

FCPs are the points in a face which can represent facial characteristics. Figure 1 shows the FCPs and a_i is a vector expressing the coordinate of FCPs. a_i is described as [7], [8]

$$a_i = (x_i, y_i),\ i = 1,2,..., 30 \tag{1}$$

To normalize the face image, we introduce a quantity, base, which is not varied for each of facial expressions, expressed as the origin of which is assigned at the top of nose, is taken for the coordinate of FCPs in this study. The information of these 30 FCPs are input to a computer by using a mouse device.

In Figure 1, X_b-Y_b coordinate system shows the absolute coordinate system and X'-Y' coordinate system is used for the new coordinate of FCPs. The origin ($origin_x$, $origin_y$) of X'-Y' coordinate system is chosen at the point of length *base* downward of the mid point between the left and right eyes [5], [6], [7], [8].

$$\mathbf{base} = \sqrt{(\mathbf{X_{b2} - X_{b1}})^2 + (\mathbf{Y_{b2} - Y_{b1}})^2} \tag{2}$$

Then we Introduce θ, which is the inclination of face with respect to the horizontal line, and defined as

$$\boldsymbol{\theta} = \mathbf{tan}^{-1}\frac{(\mathbf{Y_{b2} - Y_{b1}})}{(\mathbf{X_{b2} - X_{b1}})} \tag{3}$$

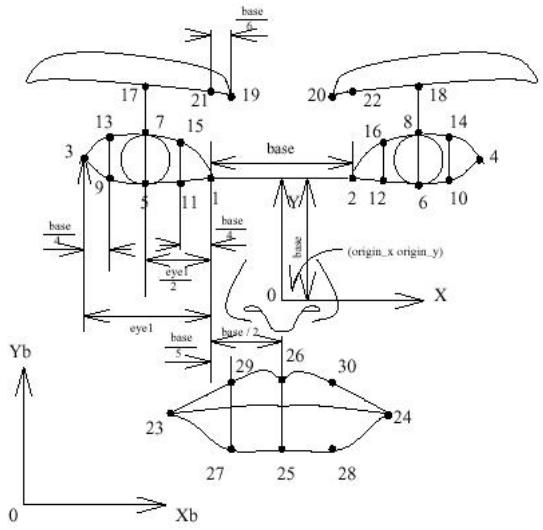


Figure 1. Facial Characteristic Points [6], [7], [8]

The X and Y coordinates, (X_0, Y_0), of the mid point between the left and right eyes are described as

$$\mathbf{X_0} = \frac{\mathbf{X_{b1} + X_{b2}}}{\mathbf{2}} \ \mathbf{and}\ \mathbf{Y_0} = \frac{\mathbf{Y_{b1} + X_{b2}}}{\mathbf{2}} \tag{4}$$

The origin of the new coordinate system is calculated as

$$origin_x = X_0 + base * \sin\theta \ \text{and}\ origin_y = Y_0 - base * \cos\theta \tag{5}$$

The coordinate (X_{bi},Y_{bi}) of a FCP a_i is transformed into X-Y coordinates system, (X_i,Y_i) by subtracting the origin coordinates and then rotating an angle of orientation θ, and their relationships are given by

$$X_i = X_{bi} - origin_x \ \text{and}\ Y_i = Y_{bi} - origin_y \tag{6}$$

$$X_{ir} = X_i \cos\theta + Y_i \sin\theta \ \text{and}\ Y_{ir} = -X_i \sin\theta + Y_i \cos\theta \tag{7}$$

We normalize the input face input face image by dividing (X_{ir},Y_{ir}) above by value base to compensate the distance effect between the client face and the camera and the size of the client face, given by

$$\mathbf{X_i'} = \frac{\mathbf{X_{ir}}}{\mathbf{base}} \ \mathbf{and}\ \mathbf{Y_i'} = \frac{\mathbf{Y_{ir}}}{\mathbf{base}} \qquad \mathbf{i = 1, 2, ..., 30} \tag{8}$$

2.2 Feature Extraction from Feature Points

Seven features were extracted from the feature vector for each expression, and were used to classify that expression to one of the six basic emotions, using RBFN. Extracted features are as follows [6], [8]:

Openness of Eyes:

$$oe = \frac{(Y_{n7} - Y_{n5}) + (Y_{n8} - X_{n6})}{2} \quad (9)$$

Width of Eyes:

$$we = \frac{(X_{n1} - X_{n3}) + (X_{n4} - X_{n2})}{2} \quad (10)$$

Height of Eyebrows 1:

$$he1 = \frac{(Y_0 - Y_{n19}) + (Y_0 - X_{n20})}{2} \quad (11)$$

Height of Eyebrows 2:

$$he2 = \frac{(Y_0 - Y_{n17}) + (Y_0 - Y_{n18})}{2} \quad (12)$$

Width of Mouth:

$$wm = X_{n24} - X_{n23} \quad (13)$$

Openness of Mouth:

$$om = Y_{n26} - Y_{n25} \quad (14)$$

Nose Tip-Lip Corners Distance:

$$nl = \frac{(X_0 - X_{n23}) + (X_0 - X_{n24})}{2} \quad (15)$$

Figure 2 are example face image of feature extraction from feature points. Standardization is needed to find out the relative displacement of the facial landmarks from their normal position [17].

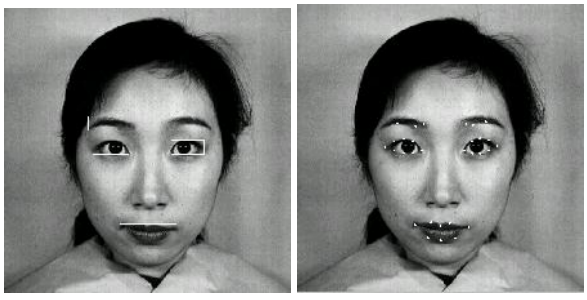


Figure 2. (a) Example of Feature Extraction, (b) Example of FCPs

3. FACIAL EMOTIONAL EXPRESSIONS

3.1 Facial Emotional Expression Synthesis

The synthesis of facial emotional expressions can be seen as a reverse process of facial emotional expressions recognition [7]. In recognition, we present the necessary information (the movements of the landmarks) so as to classify for a particular facial expression (an emotional label) in the order happy, sad, angry, fear, suprised and disgusted.

But what if we reverse the question: given a particular facial expression, can you tell what the necessary movements of the landmarks are? This indeed can be seen as a reverse process to which emotion labels are used and the outputs are the movements of the landmarks.

The JAFFE Database have 213 face images of 7 facial expressions (6 basic facial expressions and 1 neutral facial expression) taken from 10 Japanese female models. In Figure 3 are Example of Facial Emotional Expressions of JAFFE Database [17].

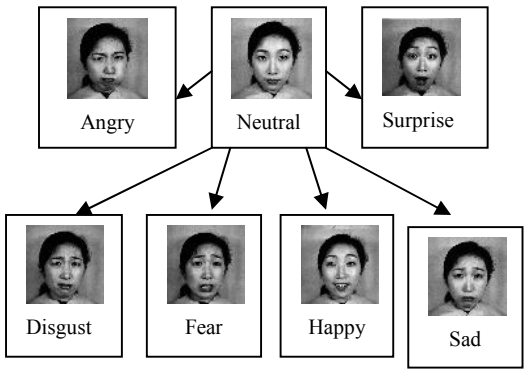


Figure 3. Example of Facial Emotional Expressions of JAFFE Database

The six basic emotions defined by [16] can be associated with a set of facial expressions. In Table 1 shows textual description of facial expressions as representations of basic emotions.

Table 1. Facial Expressions of Basic Emotions [16]

No	Basic Emotion	Textual Description of facial Expressions
1	Happy	The eyebrows are relaxed. The mouth is open and the mouth corners pulled back toward the ears.
2	Sad	The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed.
3	Fear	The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert.
4	Angry	The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose the teeth.
5	Surprise	The eyebrows are raised. The upper eyelids

		are wide open, the lower relaxed. The jaw is opened.
6	Disgust	The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often symmetrically.

3.2 Radial Basis Function Network

The basic idea is to find out the spatial differences between the FCPs of the normal face and that of the expressive face. Thus, differences of those 30 pairs of position information will constitute the 60 inputs to the two-layered neural network as Figure 4.

The basic principle of synthesizing facial expressions is to find out the necessary relative spatial shift of the FCPs for each expressions of emotion. So what is initially the input to the neural network in recognition will become the output in synthesis and vice versa [7]. RBF have proven to be an effective tool in interpolating data in multidimensional spaces.

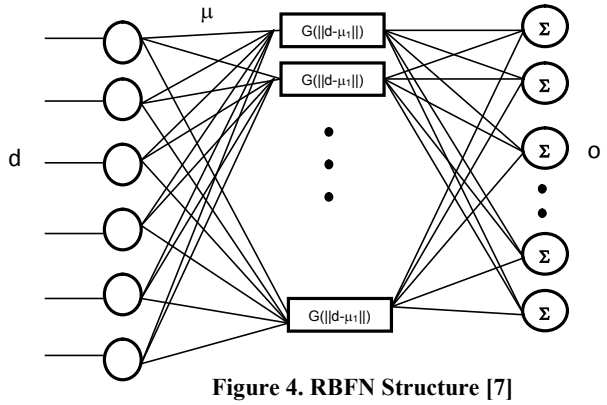


Figure 4. RBFN Structure [7]

The RBFN is ideal for interpolation since it uses a radial basis function, for example Gaussian function, for smoothing out and predict missing and inaccurate inputs [8].

We would consider interpolating functions of the form:

$$F_k(\vec{d}) = \sum_{j=1}^{m} \omega_{jk} g\left(\left\|\vec{d} - \vec{\mu}_j\right\|\right), \quad \vec{d} \in R^n, \; k = 1,..., n' \qquad (16)$$

where $\|.\|$ denotes the usual Euclidean norm on R^n and $\mu_j \in R^{n'}$, j=1, 2, ..., m denotes the centers of the radial-basis functions which are given as the known data points. Often, the g(.) is the normalized Gaussian activation function defined as

$$g(\vec{d}) = \frac{\exp\left[-(\vec{d} - \mu_j)^2 / 2\sigma_j^2\right]}{\sum_k \exp\left[-(\vec{d} - \mu_k)^2 / 2\sigma_k^2\right]} \qquad (17)$$

where d is the input vector, μ is a set of weights and σ is the width of the RBF.

Hence, the determination of the nonlinear map F(*d*) has been reduced to the problem of solving the following set of linear equations for the coefficients ω_j,

$$\begin{pmatrix} f_{1k} \\ \vdots \\ f_{mk} \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mm} \end{pmatrix} \cdot \begin{pmatrix} \omega_{1k} \\ \vdots \\ \omega_{mk} \end{pmatrix}, \; k=1,2, ..., n' \qquad (18)$$

where $A_{ij} = g\left(\left\|\vec{d} - \vec{\mu}_j\right\|\right), i, j = 1, 2, ..., m.$

RBFN is class of single hidden layer feedforward networks where the activation functions for hidden units are defined as radially symmetric basis functions phi such as the Gaussian function. The fraction of overlap between each hidden unit and its neighbors is decided by width sigma such that a smooth interpolation over the input space is allowed. The whole architecture is therefore fixed by determing the hidden layer and the weights between the middle and the output layers.

The number of input layer units must be equal to 7, equal to the number of extracted features, and that of output layers is 6, which corresponds to six kinds of facial expressions. The network training is carried out by back propagation algorithm.

3.3 FEES Main Process using RBFN

FEES main process using RBFN consists of three processes, namely sampling, training and testing. In the process of sampling, we collected data on FCPs displacement of the images are entered. In the process of training, we train the RBFN to identify FCP displacement for various expressions. While in the process of testing, we use the knowledge possessed by the RBFN to perform FEES in accordance with the desired expression. The main processes in this paper can be seen in Figure 5.

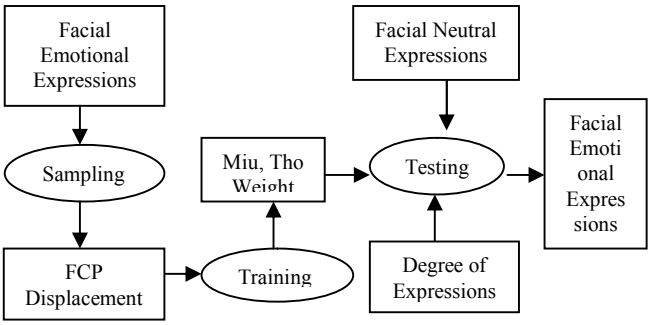


Figure 5. FEES Main Process using RBFN

4. EXPERIMENTAL RESULT

A set of 256 x 256 grayscale images are used in our experiment. In the RBFN classifier for FEES, we used 6 input layer, 10 hidden layer, and 60 output in 60 sample. We just did a bit of input samples and hidden units to be more easily studied. The more input samples and the number of hidden units, then the result will be better.

Learning rate used is 0.01. The process is divided into two parts, namely the process of calculating the hidden layer by using k-means clustering and the training process the input. The total time for facial feature extraction, pre-processing, neural network calculation takes less than 15 seconds.

Table 2. RBF Classifier Test Result

Basic Emotion	Degree of Expression						Result %
	Ha	Sa	Fe	An	Su	Di	
Happy	4.73	1.32	1.25	1.27	1.28	1.26	94.60
Sad	1.30	4.65	1.93	1.90	1.00	2.52	93.00
Fear	1.21	2.85	4.47	2.17	3.49	3.17	89.40
Angry	1.37	2.13	1.78	4.64	1.59	2.91	92.80
Surprise	1.81	1.49	2.54	1.62	4.24	1.00	84.80
Disgust	1.18	2.43	2.41	2.66	2.01	4.74	94.80
Average							91.57

Ha=Happy,Sa=Sad,Fe=Fear,An=Angry,Su=Suprise,Di=Disgust

Table 2 shows that the high value of the Degree of Expression equivalent to the Basic Emotion. This percentage impact on the results obtained by comparing the maximum value of expression levels. So that the average of the RBF Classifier Test Result in this paper is 91.57%.

5. CONCLUSION

In this research we presented two systems for classifying of the facial expressions from JAFFE Database. In the RBFN classifier, 7 feature extracted from 30 feature points were used as training and test sequences. The trained RBFN was tested by features that not used in training and we have obtained a high result rate of 91.57%.

6. REFERENCES

[1] M.S. El-Nasr, "Modeling Emotion Dynamics in Intelligent Agents," M.S. Thesis, May 1988.

[2] Paul Ekman and W.V. Friesen, "Facial Action Coding System (FACS)," Consulting Psychologists Press, Inc., 1978.

[3] Paul Ekman, "Facial Expressions of Emotion: New Findings, New Questions," American Psychological Society, Vol. 3, No. 1, pp. 34-38, January 1992.

[4] Paul Ekman, "Facial Expressions and Emotion," American Psychological, Vol. 48, pp. 384-392, 1993.

[5] Hiroshi Kobayashi and Fumio Hara, "The Recognition of Basic Facial Expressions by Neural Network," in International Joint Conference on Neural Network, pp. 460-466, 1991.

[6] Hiroshi Kobayashi and Fumio Hara, "Recognition of Six Basic Facial Expressions and Their Strength by Neural Network," IEEE International Workshop on Robot and Human Communication, pp. 381-386, 1992.

[7] Mary Y.Y. Leung, Hung Yen Hui and Irwin King, "Facial Expression Synthesis by Radial Basis Function Network and Image Warping," In IEEE International Conference on Neural Networks, Vol III, pp. 1400-1405, Washington D.C., 1996. IEEE Computer Society.

[8] I. King and H.T. Hou, "Radial Basis Function for Facial Expression Synthesis," In IEEE International Conference on Neural Networks, 1996.

[9] J. Yau and A. Duffy, "A Texture mapping approach to 3D facial image synthesis, 1988.

[10] Fumio Hara and Hiroshi Kobayashi, "Computer Graphic for Expressing Robot-Artificial Emotions," IEEE International Workshop on Robot and Human Communication, pp. 381-386, 1992.

[11] D.R. Nur Arad, Nira Dyn and Y. Yeshurun. Image Warping by Radial Basis Functions: Application to facial Expressions. CVGIP: Graphical Models and Image Processing, 56, No. 2:161-172, 1994.

[12] P. Ekman and W.V. Friesen, "Unmasking the Face," Consulting Psychologists Press, Inc., 1975.

[13] Jyh-Yeong Chang and Jia-Lin Chen, "A Facial Expression Recognition System Using Neural Networks," IEEE, pp. 3511-3516, 1999.

[14] Daw-Tung Lin and Jam Chen, "Facial Expressions Classification with Hierarchical Radial Basis Function Networks," IEEE, pp. 1202-1207, 1999.

[15] Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi, "Recognition of Basic Facial Expressions by Feature-Points Tracking using RBF Neural Network and Fuzzy Inference System," IEEE International Conference on Multimedia and Expo (ICME), pp. 1219-1222, 2004.

[16] Surya Sumpeno, Mochamad Hariadi, and Mauridhi Hery Purnomo, "Facial Emotional Expressions of Life-like Character Based on Text Classifier and Fuzzy Logic," in IAENG International Journal of Computer Science, 38:2, IJCS_38_2_04 [online], May 2011.

[17] The Japanese Female Facial Expression (JAFFE) Database. Available: [Online] http://www.ksrl.org/jaffe.html.

[18] Muid Mufti, Assia Khanam, "Fuzzy Rule Based Facial Expression Recognition," IEEE Computer Society, International Conference on Computational Intelligence for Modelling Control and Automation, and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), 2006.

[19] Hadi Seyedarabi, Ali Aghagolzadeh, and Sohrab Khanmohammadi, "Facial Expression Recognition from Static Images using RBF Neural Networks and Fuzzy Logic," 2nd Iranian Conference on Machine Vision and Image Processing (MVIP 2003), Tehran, 2003.

[20] David Matsumoto and Bob Willingham, "Spontaneous Facial Expressions of Congenitally and Noncongenitally Blind Individuals," Journal of Personality and Social Psychology, 2009 American Psychological Association, Vol. 96, No. 1, pp. 1-10, 2009.

[21] Keith Anderson and Peter W. McOwan, "A Real-Time Automated System for the Recognition of Human Facial Expressions," IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, Vol. 36, No. 1, February 2006.

[22] Marian Stewart Bartlett, Gwen Littlewort, Claudia Lainscsek, Ian Fasel, Javier Movellan, "Machine Learning Methods for Fully Automatic Recognition of Facial Expressions and Facial Actions," IEEE International Conference on Systems, Man and Cybernetics, pp. 592-597, 2004.

[23] R. Ravi and Mahima S, "Study of the Changing Trends in Facial Expression Recognition," International Journal of Computer Application (0975-8887) Vol. 21, No. 5, pp. 10-16, May 2011.

[24] Spiros V. Ioannou, Amarylis T. Raouzaiou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzies, Stefanos D. Kollias, "Emotion Recognition through Facial Expression Analysis based on a Neurofuzzy Network," Elsevier Ltd. All rights reserved. doi:10.1016/j.neunet.2005.03.004. Neural Networks 18 (2005), pp. 423-435, 2005, Available online at www.sciencedirect.com.

[25] V. Radha, and N. Nallammal, "Neural Network Based Face Recognition Using RBFN Classifier," Proceedings of the World Congress on Engineering and Computer Science, WCECS 2011 in San Francisco, USA, Vol. I, October 19-21, 2011.

Indonesian Vehicle Plate Recognition and Identification Based on Digital Image Processing and Artificial Neural Network

Yuli Sun Hariyani
Faculty of Electrical & Communication
IT Telkom
Bandung, Indonesia
yoeli.sun@gmail.com

Inung Wijayanto
Faculty of Electrical & Communication
IT Telkom
Bandung, Indonesia
inung_wijayanto@yahoo.com

ABSTRACT
Every vehicle has its own license which is given legally. In Indonesia, vehicle registration number is an important thing used by system such as: parking system, building security system and toll system.

This research is doing a vehicle plate recognition and identification which will read the characters on the plate. The input comes from a real time video. There are six main processes in this system. In preprocessing, the system will enhance frame by using top hat and bottom hat transformation. While in detection, it will detect and crop the plate by using integral protection. After getting the plate position, an identification process will do by doing segmentation in each character. Then a feature extraction which combined with Artificial Neural Network is done to get the character of the vehicle registration number.

Keywords
Vehicle Registration Number, Self Organizing Maps

1. INTRODUCTION
Every vehicle has its own identities; one of them is a plate with vehicle registration number. The identity mostly used in parking system, building security system and toll system. In order to make the system reliable, it is needed to develop an automatic system which can recognize and identify the vehicle registration number.

Every nation has their own license plate standard and it's different between one and another. For Indonesia, standard of license plate is black background with white character. Several research to detecting and extracting plate are using bounding box method [11] or Hough transform [6] that detect the line of background plate to know where the plate is, but those kind of methods will fail when it's used for detecting and extracting Indonesian license plate from a black car.

This research proposes combination method to detect and extract license plate that suitable for Indonesian license plate. And for the accuracy adjustment the system divide the plate into three parts: the first and the third plate are alphabet, and the second plate is number. By doing this, the identifying between alphabet and number are done separately and this method will reduce the possibility of wrong detection between the similar pattern of alphabet and number such as O and 0.

This research proposes combination method to detect and extract license plate that suitable for Indonesian license plate. And for the accuracy adjustment the system divide the plate into three parts: the first and the third plate are alphabet, and the second plate is number. By doing this, the identifying between alphabet and number are done separately and this method will reduce the possibility of wrong detection between the similar pattern of alphabet and number such as O and 0.

This research proposes combination method to detect and extract license plate that suitable for Indonesian license plate. And for the accuracy adjustment the system divide the plate into three parts: the first and the third plate are alphabet, and the second plate is number. By doing this, the identifying between alphabet and number are done separately and this method will reduce the possibility of wrong detection between the similar pattern of alphabet and number such as O and 0.

The main objective of this research is to develop a system that can detect and identify the number of from vehicle registration plate.

2. SYSTEM DESIGN
2.1 VIDEO INPUT
The input comes from a video got from a digital camera with 640x480 pixel using *.avi format. Then an extraction has done by arranging the frame as a digital image.

2.2 PREPROCESSING
Each frame from extraction is processed independently. RGB conversion to grayscale, frame cutting, top-hat and bottom-hat transformations. Frame cutting is done with an assumption that the vehicle plate is on the lower location of the frame.

Top-hat and bottom-hat transformation is a transformation based on mathematic morphology. This transformation will expand the grey region so that the region which the vehicle plate located will clearly visible.

Figure 1: RGB image

2.3 PLATE DETECTION

After preprocessing, an edge detection and low pass filtering is done to eliminate the noise in the image. The detection of the vehicle plate is done by using integral projection method. This method is capable to find the region of the object by adding the pixel each row and column.

The plate region has more pixel than the other region. The threshold value is the average number of pixel from the image from edge detection process.

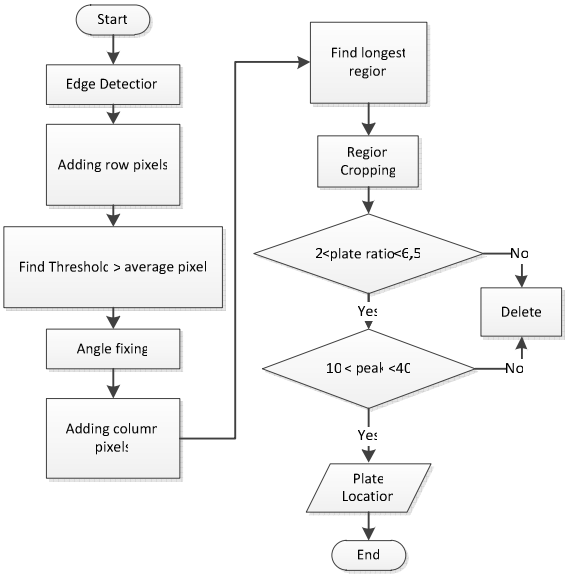


Figure 2: Plate Detection Flowchart

2.4 SEGMENTATION AND NORMALISATION

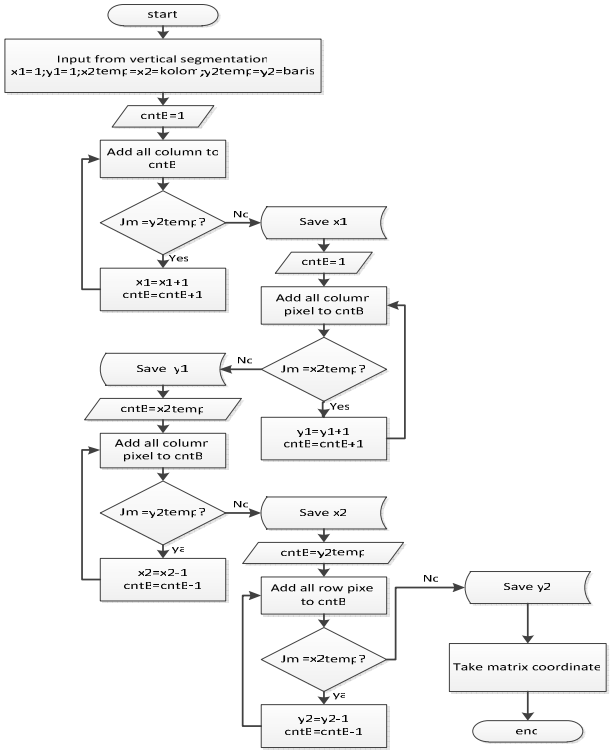


Figure 3: Edge segmentation

Based on the characteristic of the vehicle plate in Indonesia, the region of the characters is divided into three parts. This is done in order to increase the accuracy of the character recognition.

There are three segmentations done in this process, horizontal segmentation, vertical segmentation and edge segmentation. Basically the horizontal and vertical segmentation are similar, the different is in the iteration process. In horizontal segmentation, iteration is done from the first row, while in vertical segmentation the iteration is done from the first column.

Edge segmentation algorithm:
1. Inverse the image's color from vertical segmentation.
2. Adding row pixels into one column.
3. while total pixel = total row, do next column iteration; while total pixel = total row stop iteration
4. Save coordinate, repeat step 2 and 3.

2.5 FEATURE EXTRACTION

Feature extraction used in this research is vector based feature extraction. Each 8x8 pixels matrix average is being calculate then obtained a 1x8 pixels. The average pixels are done by summed every 8x8 pixels.

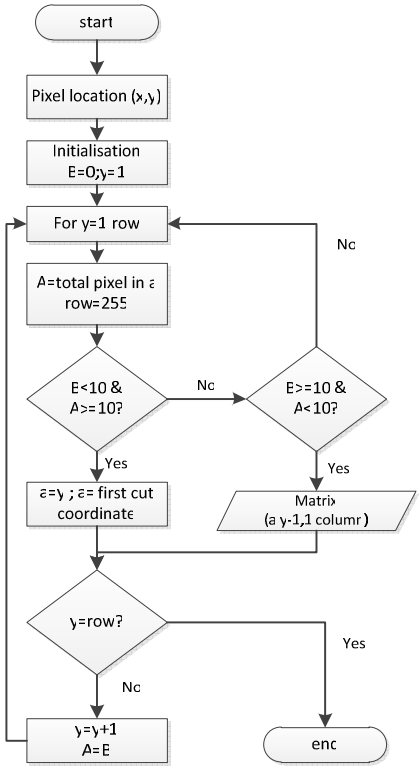


Figure 4: Horizontal Segmentation

2.6 Character Recognition using Self Organizing Maps

The input of this process comes from feature vectors of each character. The output this process are 26 classes of character and 10 classes of number. Because the input vector's size is 1x8 matrixes, the neural network build eight neuron input.

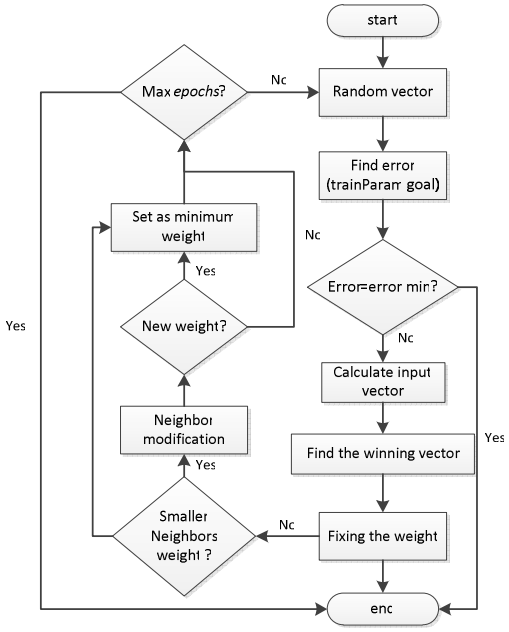


Figure 5: Self Organizing Maps

Self-organizing maps (SOM) learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on.

The neurons in the layer of an SOM are arranged originally in physical positions according to a topology function. The function gridtop, hextop, or randtop can arrange the neurons in a grid, hexagonal, or random topology. Distances between neurons are calculated from their positions with a distance function. There are four distance functions, Euclidean distance, box distance, link distance, and Manhattan distance.

$$dj = \sum_{i=0}^{n-1} (xi(t) - wij(t))^2$$

In learning process, the input is an 8x26 matrix for character and 8x9 for number. Every character is got from the feature extraction method. The matrix is ordered in a row so that the final matrix is 28x8 and 9x8. Next step is transposing the matrix into 8x26 and 8x9 as input to neural network. The transpose is done because the class of the neural network is done by column.

2.7 System Testing

The testing is done for each different parameter of SOM neural network. So, analysis of the accuration can be done by seeing the effect of each parameter. Systematically the calculation of the accuration can be written as:

$$Accurations = \frac{total\_correct\_result}{total\_data} x100\%$$

3. EXPERIMENT RESULT

Experiments have been performed to test the proposed system.

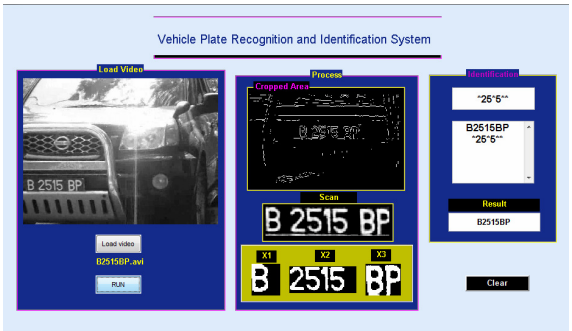


Figure 6: User Interface

Table 1: Character identification

Topology	Distance Function	Total character	Correct character	Missed character	Accuration
Gridtop	Dist	137	119	18	86,86%
	mandist	137	122	15	89,05%
	linkdist	137	120	17	87,59%
	boxdist	137	121	16	88,32%
Hextop	dist	137	121	16	88,32%
	mandist	137	118	19	86,13%
	linkdist	137	122	15	89,05%
	boxdist	137	134	3	97,81%
Randtop	dist	137	120	17	87,59%
	mandist	137	122	15	89,05%
	linkdist	137	119	18	86,86%
	boxdist	137	117	20	85,40%

From table 1 can be seen that the best accuration of the detection using combination of the hexagonal topology and box distance function. The accuration is 97,81%.

Figure 7: Image output from edge detection

Table 2: Distance function testing

Topologi	*Epochs*	*Learning Rate* (α)	Fungsi Jenis *Distance*	Accuration
Gridtop	200	0.9	*Linkdist*	87.35%
Hextop	200	0.9		76.82%
Randtop	200	0.9		80.92%
Gridtop	200	0.9	*Dist*	86.67%
Hextop	200	0.9		86.34%
Randtop	200	0.9		85.36%
Gridtop	200	0.9	*Mandist*	81.16%
Hextop	200	0.9		88.30%
Randtop	200	0.9		85.09%
Gridtop	200	0.9	*Boxdist*	83.42%
Hextop	200	0.9		89.23%
Randtop	200	0.9		86.67%

By using the different pairs of topology and function distance, the system can be obtained the different result. The system is tested by using 20 videos that contain license plate and complex background which is taken around 08.00 – 11.00 am.

Figure 8: Plate Cropping

From the result in table 1, it can be concluded that the best result is by using hextop topology and boxdist for the distance function. While in general identification, there are 3 mistakes that made from 20 input.

4. CONCLUSION

The proposed system for the detecting and extracting plate is reliable for the 20 video testing which is has complex background and the system can detect the plate even though there are no assumptions of distance between plate and camera.

The best accuracy for character identification is obtained by using hexagonal topology and box distance with 97, 81 % in average.

5. REFERENCES

[1] Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J. 1996.SOM_PAK: The self-organizing map program package. Report A31.Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland. Also available in the Internet at the address http://www.cis.hut.fi/research/som_lvq_pak.shtml.

[2] Gonzales, Rafael C. and Woods, Richard. 1993. *Digital Image Processing*.USA: Addison-Wesley Publishing Company.

[3] Hou, P G., J Zhao and M Liu. 2006. *A License Plate Locating Method Based on Tophat-bothat Changing and*

Line Scanning. Journal of Physics: Conference Series, 48, 431-436.

[4] Kenneth R. Castleman. 1996. *Digital Image Processing*.Prentice Hall

[5] Pandjaitan, L.W. 2007. *Dasar-dasar Komputasi Cerdas*.Yogyakarta:Andi Offset.

[6] Saha, Satadal, et al. 2009. *License Plate Localization from Vehicle Images: An Edge Based Multi-stage Approach.* International Journal of Recent Trends in Engineering, Vol 1, No. 1

[7] Siang, Jong Jek. 2005. *Jaringan Syaraf Tiruan & Pemrogramannya Menggunakan MATLAB*. Yogyakarta:Andi.

[8] Sukmawan, Arief. 2008. *Implementasi Pembacaan Huruf Hijaiyyah dengan Pendekatan Struktur Melalui Ekstraksi Ciri Vektor dan Pengenalan Menggunakan Jaringan Syaraf Tiruan SOMs (Self-Organizing Maps)*.Jurusan Teknik Telekomunikasi IT Telkom Bandung.

[9] Suyanto. 2007. *Artificial Intelligence*.Bandung:Informatika.

[10] Wijaya, M.Ch. and Prijono, A. 2007. *Pengolahan Citra Digital Menggunakan Matlab*.Bandung: Informatika.

[11] Wijayanto, I. 2008. *Identifikasi Nomor Polisi Kendaraan Bermotor Menggunakan Jaringan Syaraf Tiruan Self Organizing Maps(SOMs)*.Jurusan Teknik Telekomunikasi IT Telkom Bandung.

[12] William K.P. 1991. *Digital Image Processing*.Wiley-Interscience Publication.

[13] Yanamura, Y. 2003. Extraction and Tracking of the License Plate Using Hough Transform and Voted Block Matching. Intelegent Vehicles Symposium. Proceedings. IEEE.

Wavelet Types Comparison for Extracting Iris Features Based on Energy Compaction

R. Rizal Isnanto
Doctoral Student at Electrical Engineering, Gadjah Mada University
Jl. Grafika No. 2 Yogyakarta
+62247477182; +628122731568
rizal_isnanto@yahoo.com

Suhardjo
Ophthalmology Dept., Faculty of Medicine, Gadjah Mada University
Jl. Farmako, Sekip Utara, Bulaksumur, Yogyakarta
+62274560300
suhardjo2005@yahoo.com

Thomas Sri Widodo
Electrical Engineering and Information Technology Dept.,
Gadjah Mada University
Jl. Grafika No. 2 Yogyakarta
+62274552305
thomas@te.ugm.ac.id

Adhi Susanto
Electrical Engineering and Information Technology Dept.,
Gadjah Mada University
Jl. Grafika No. 2 Yogyakarta
+62274552305
susanto@te.ugm.ac.id

ABSTRACT
Human iris has a very unique pattern which is possible tobe used as a biometric recognition. To identify texture in an image, texture analysis method can be used. One of method is wavelet that extract the image feature based on energy. Wavelet transforms used are Haar, Daubechies, Coiflets, Symlets, and Biorthogonal. In the research, iris recognition based on five mentioned wavelets was done and then comparison analysis was conducted for which some conclusions taken. Some steps have to be done in the research. First, the iris image is segmented from eye image then enhanced with histogram equalization. The features obtained is energy value. The next step is recognition using normalized Euclidean distance. Comparison analysis is done based on recognition rate percentage with two samples stored in database for reference images. After finding the recognition rate, some tests are conducted using Energy Compaction for all five types of wavelets above. As the result, the highest recognition rate is achieved using Haar, whereas for coefficients cutting for C(i) < 0.1, Haar wavelet has a highest percentage, therefore the retention rate or significan coefficient retained for Haar is lower than other wavelet types (db5, coif3, sym4, and bior2.4)

Keywords
Energy compaction, recognition rate, wavelets coefficients, retention.

1. INTRODUCTION

1.1 Background
Humans as individuals, have unique characteristics and distinctive. These characteristics can be used to recognize or identify persons. This is known as biometric recognition. Iris is the part of the circle around eye pupil. Although iris has a relatively narrow region compared with entire area of the human body, iris has a very unique pattern, different in each individual and the pattern will remain stable. For those reasons, iris can be used as the basis for the recognition in biometrics.

Many algorithms have been applied as a method of iris recognition, such as PCA (Principal Component Analysis), ICA (Independent Component Analysis), Gabor-Wavelet algorithm [17], characterizing Key Local Variation, Laplace Pyramid, Gray Level Co-ocurrence Matrix (GLCM) [8] and others. Haar wavelet transform as a method to analyze the texture is still rarely used as a feature extractor on iris pattern. In this study, texture-based recognition methods were analyzed using the method of characterizing the transforms of Haar, Daubechies, Coiflets, Symlets, and Biorthogonal wavelets, which then the comparison of these five types of wavelets based on energy compaction will be conducted.

1.2 Research Objectives
The purpose of doing this research is to analyze some types of wavelets that can perform a comparative analysis on iris recognition rate using the Haar, Daubechies (db5), Coifles (coif3), Symlets (sym4), and Biorthogonal (bior2.4) wavelets with Energy Compaction calculations. Therefore those results then are compared and analyzed against their recognition rates. While the motivation of the research is that the iris can be used as an organ in biometric system with higher accuracy level and the fact that, up until now, there is no adequate research done for comparing these five wavelets mentioned above to extract iris features based on iris compaction.

1.3 Limitations
In order not to deviate far from the problems exist, the research problems are limited as follows:
1. Iris image used is the image that has been available, which is using the database of CASIA V1.0.
2. Research is focused on the compaction energy of Haar Daubechies (db5), Coifles (coif3), Symlets (sym4), and Biorthogonal (bior2.4) wavelets transform as feature extractor and feature used is energy
3. Recognition method used is the normalized Euclidean distance method

2. BASIS OF THEORY

2.1 Iris Eyes

Iris can serve as the basis for biometric systems. Each iris has a texture that is very detailed and unique to each person and remain stable for decades. The eye can not be altered through surgery without causing any damage to eyesight. Figure 1 shows the anatomy of the eye, and examples of human iris[12].

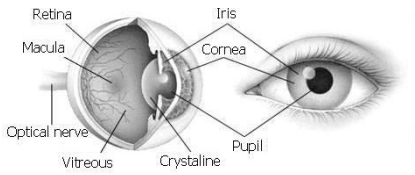


Figure 1. Anatomy of eyes and example of iris region.

The advantages of using the iris for reliable identification system are [12] as follows.

1. Iris is insulated and shielded from the outside environment.
2. On the iris, it is not possible to do some surgeries without causing defects in the eye.
3. Iris has a physiological response to light, which allows testing of the natural use of the possibility of fraud and faked eye lenses and so forth.

2.2 Haar Wavelet

First type of wavelet used in this study is the Haar wavelet. Haar wavelet is a compactly supported wavelets, wavelets of the oldest and simplest. Haar wavelet is in category of orthogonal. Haar wavelet is also similar with db1 wavelet (Daubechies order 1). Haar wavelet filter length is 2. Haar Wavelet scaling function is shown in Figure 2.

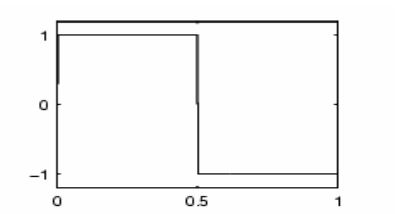


Figure 2. Scaling function of Haar wavelet

2.3 Daubechies Wavelets

Daubechies wavelet has a short name db, and to order N denoted by db$N$. For order of Daubechies Wavelet $N = 1$ is also called Haar, Order range of Daubechies wavelet is from $N = 2$ until $N = 45$. Daubechies wavelet length is $2N$. For example db5 has a filter length 10.

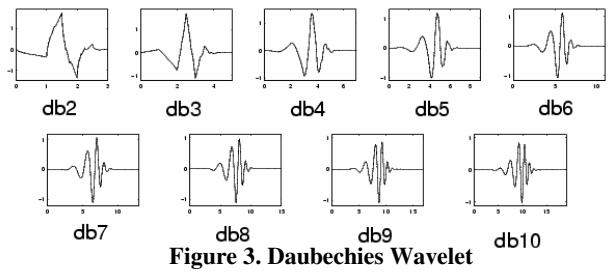


Figure 3. Daubechies Wavelet

Figure 3 shows the length of several types of wavelet filters of Daubechies from $N = 2$ to $N = 10$.

2.4 Coiflet Wavelets

Coiflet wavelets have a short name coif, and to order N *denoted by* coif$N$. Coiflet wavelets length is 6N. Figure 4 shows the filter lengths of Coiflet wavelets from $N = 1$ sampai $N = 5$.

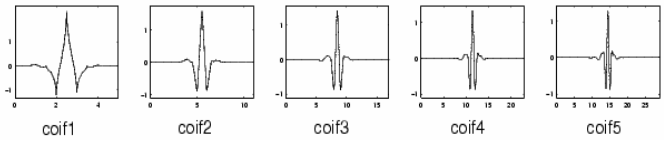


Figure 4. Coiflet Wavelet.

2.5 Wavelet Symlet

Symlets are orthogonal and compactly supported wavelets, which introduced first by I. Daubechies as a modification of db family wavelets. Symlets have near symmetry property and have a leat asymmetrical level. [20]. The symlet wavelets and scaling functions for orders 2 up to 8 are shown in respectively Figure 5 up to Figure 11.

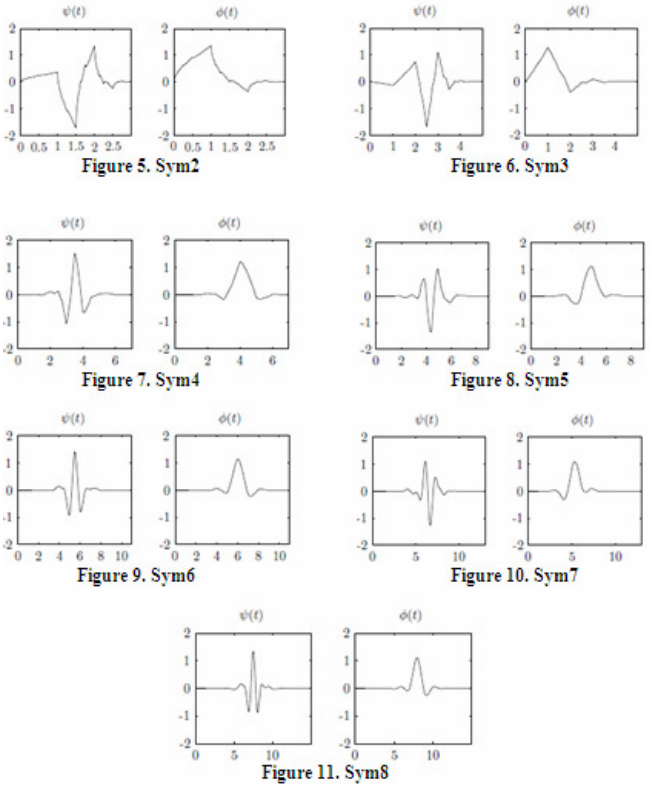


Figure 5-11. Symlet wavelets from order 2 up to order 8.

2.6 Biorthogonal Wavelets

The biorthogonal family uses separate wavelet and scaling functions for the analysis and synthesis of a signal. The reverse biorthogonal family uses the synthesis functions for the analysis and vice versa. Figure 9 up to Figure 12 show some examples of biorthogonal analysis and synthesis functions. The double

numbers in the names denote the orders of respectively the analysis and synthesis filters.

2.7 Energy Compaction

Energy is defined as the sum of the squares of the values. So the energy of an image is the sum of the squares of the pixel values, the energy in the wavelet transform of an image is the sum of the squares of the transform coefficients. During wavelet analysis the energy of a signal is divided between approximation and details signals but the total energy does not change. During compression, however, energy is lost because thresholding changes the coefficient values and hence the compressed version contains less energy.

The compaction of energy describes how much energy has been compacted into the approximation signal during wavelet analysis. Compaction will occur wherever the magnitudes of the detail coefficients are significantly smaller than those of the approximation coefficients. Compaction is important when compressing signals because the more energy that has been compacted into the approximation signal the less energy can be lost during compression.

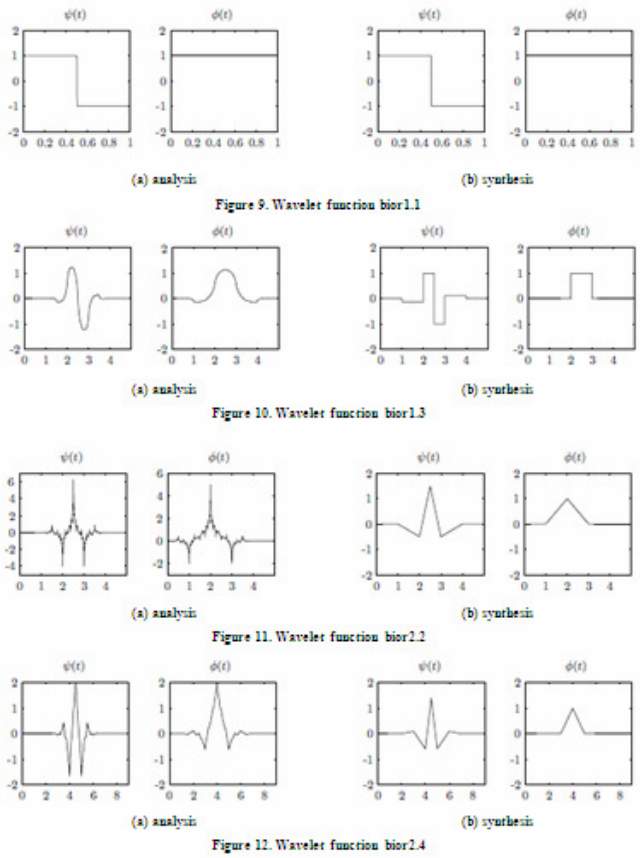


Figure 9-12. Some biorthogonal wavelet functions.

Energy compaction can be obtained from the number of least significant coefficients, for which those coefficients can be cut or removed. In this case, it can be said that insignificant coefficients mean that there are zero valued coefficients and which can be seemed less enough to be cut or removed. In energy compaction, the term of 'retention' is also used. Retention indicates coefficient values which can be retained in calculation.

3. DESIGN OF RESEARCH

3.1 Flowchart of research

Flow of iris recognition system can be depicted in a flowchart which is shown in Figure 13. Broadly explained, the processes can be classified into some steps below.

1) Selecting the input image in the form of eye image.
2) Segmenting the iris for separating iris image from its eye image.
3) Normalizing the iris for getting iris image which is easier to be manipulated.
4) Extracting the features using five types of wavelets: Haar, db5, coif3, sym4, and bior2.4.
5) Recognizing using Euclidean distance calculation.
6) Doing the 4-th step using five different wavelet types.
7) Comparing the recognition between feature extraction using all 5 types of wavelets mentioned in no. 4.
8) Taking the conclusions of point 7.
9) Plotting graphics of correlation between the number of coefficient (x-axis) against coefficients values (y-axis) for five different wavelet mentioned before.
10) Calculating the energy compaction with some value $C(i)$ which may cause different results of retention for all 5 types of wavelets.
11) Comparing and making conclusion point 8 against point 10.

3.2 Software Design

3.2.1 Iris Image Segmentation

In database, pupil circle and iris circle are not always perfect circle for which for getting only iris needs a complex computation. For simplified computation, it is assumed that pupil as well as iris have a perfect-circle form. First step is finding pupil circle, center point and its radius. Some subprocesses have to be done are thresholding, smoothing, and obtaining the center point of pupil and its radius using Circular Hough Transform.

3.2.2 Conversion into Rectangular Image

Pupil and iris images which are circle-form with their diameters are varied, are then converted into a fix-sized rectangular form. The size is 60×512 pixels. This conversion also for simplifying both coding as well as computation.

The presence of eyelids and eyelashes also can can disrupt the process of iris recognition and reduce the level of accuracy. Therefore, not all parts of the iris is taken. For the iris image of CASIA, the top part of the image is cut so that the bottom of the iris of the eye are taken. Image size is 60 × 384 pixels.

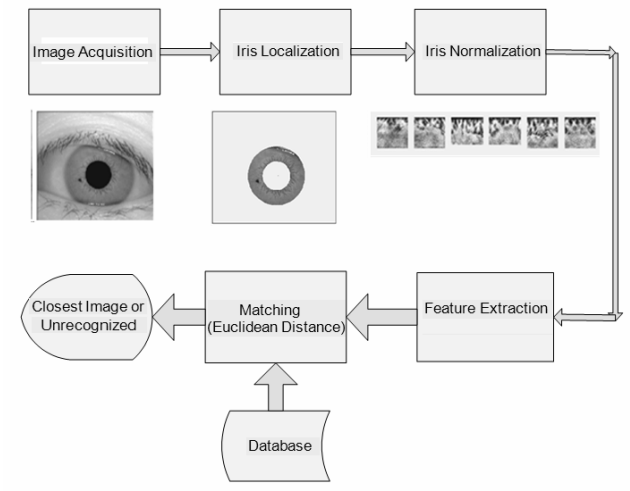


Figure 13. Flowchart of iris recognition system using feature extraction from some different types of wavelet.

3.2.3 Image Enhancement

Iris image which has been converted into a rectangular shape has a low contrast level so that the resulting level of accuracy is less good. Therefore, the contrast then enhanced using adaptive histogram equalization. First, image is is split into 3 × 12 parts of each section measuring 20 × 32 pixels. The purpose is to obtain images with good contrast but will not damage the overall image quality.

3.2.4 Feature Extraction

Iris image is split into 6 slices. Feature extraction is per-formed to all images which have been split before using 5 wave-let types. Features obtained are energies of E_a, E_h, E_v, dan E_d.

4. RESULTS AND DISCUSSION

4.1 Recognition Testing

Some tests were conducted for 128 images, originated from 64 persons. While the stored images in database are 2 samples for each person. Maximum decomposition levels for Haar, db5, coif3, sym4, and bior2.4 are 5, 2, 1, 3, and 2, respectively. From all allowed decomposition level tests, it can be inferred that: (a) The best recognition level is 84.375% which is achieved by Haar with decomposition level 3. (b) The decreasing order of average of recognition rate is: Haar, sym4, db5, and bior2.4 for the least. Table 1 shows all wavelet tests up to the maximum level achieved. From this table, it can be shown that the higher level of decomposition, the recognition level tends to be better.

Table 1. Recognition rate from different decomposition level.

Decomposition Level	Recognition Rate (%)				
	Haar	db5	coif3	sym4	bior2.4
1	68.750	38.231	25.000	42.188	42.969
2	76.563	68.750	-	60.938	60.156
3	84.375	-	-	65.625	-
4	83.594	-	-	-	-
5	83.594	-	-	-	-
Average of Recognition Rate (%)	**79.375**	**53.491**	**25.000**	**56.250**	**51.563**

4.2 Energy Compaction Testing

In this test, wavelet coefficient threshold $C(i)$ of 0.001 and 0.1 respectively are applied. We can use 2 (two) scenarios to analyze the comparison of energy compaction for 5 types of wavelet tested. *First*, we can visually analyze from graphics which depicting the relationship between i-th coefficient and its corresponding wavelet coefficient value $C(i)$. Figures 14-18 show these relationships and we therefore can observe the existing trends of the graphics. It can be shown that wavelet coefficient distribution of Haar wavelet gives the best retention because of most coefficients have a small value, and the first values are very high. It means that the approximation values are high enough if compared with detail values.

Second, we can use the retention concept which means coefficient values which can be retained in calculation. The more coefficients can be cut or removed, the less retention can be produced. The retention is better when its value is lower. Table 2 and Table 3 depicts the retention level for threshold value 0.001 and 0.1, respectively.

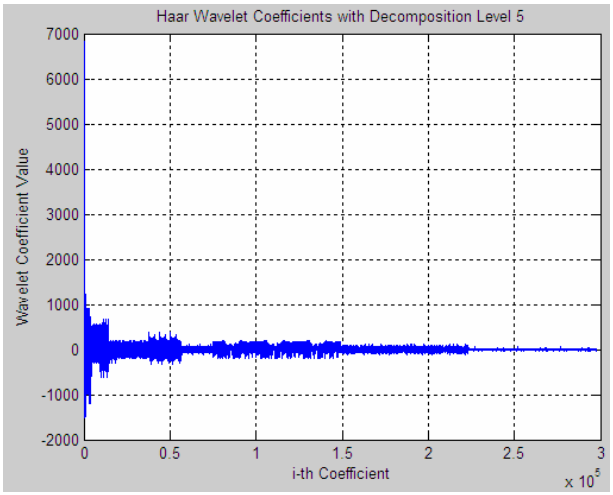


Figure 14. Wavelet coefficient values for Haar, decomposition level 5.

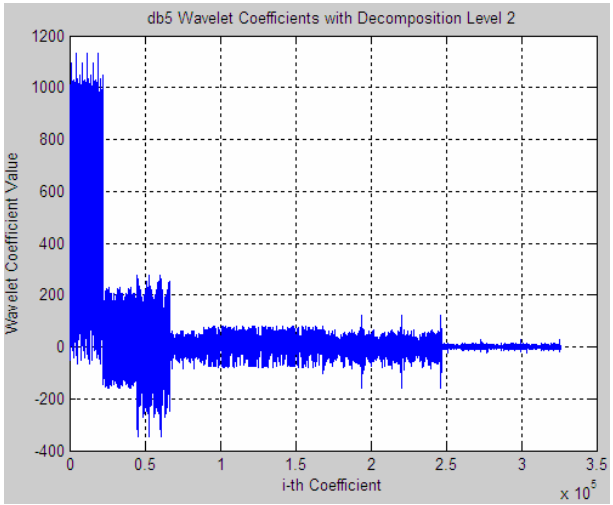


Figure 15. Wavelet coefficient values for db5, decomposition level 2.

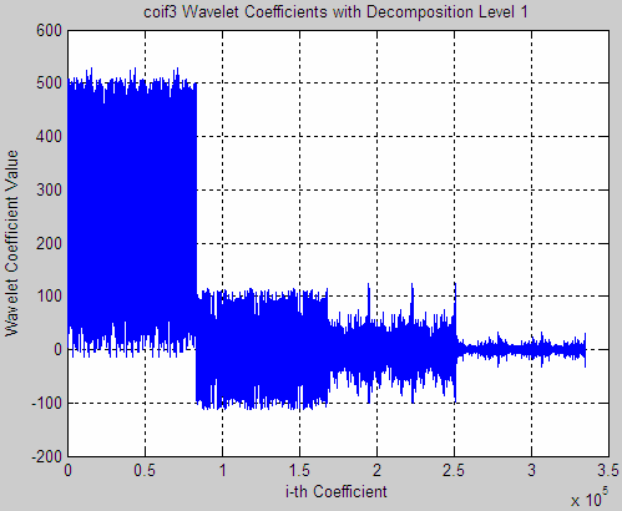


Figure 16. Wavelet coefficient values for coif3, decomposition level 1.

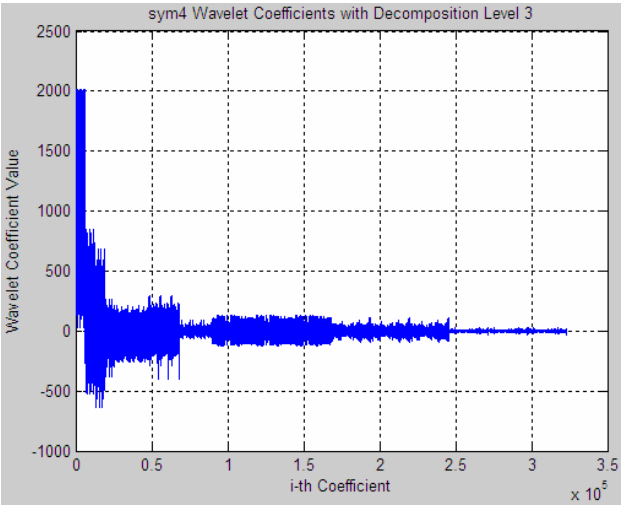


Figure 17. Wavelet coefficient values for sym4, decomposition level 3.

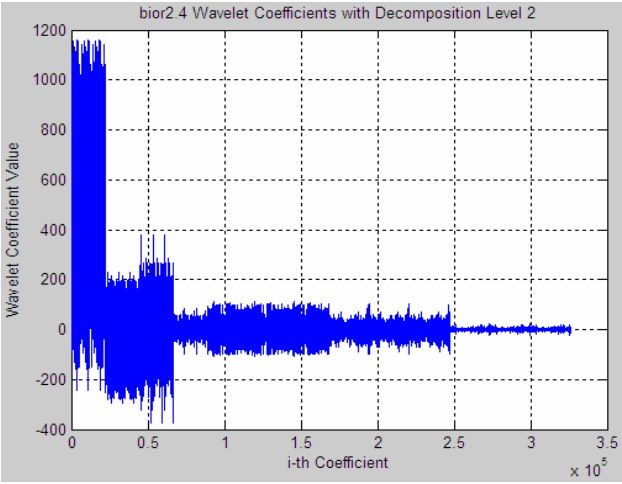


Figure 18. Wavelet coefficient values for bior2.4, decomposition level 2.

Table 2. Energy Compaction with Threshold C(i) > = 0.001

Wavelet	Energy compaction	Decomposition Level				
		1	2	3	4	5
Haar	Retention (%)	58.7	49.3	46.8	46.3	46.1
	Removed (%)	41.3	50.7	53.2	53.7	53.1
db5	Retention (%)	63.1	54.2	-	-	-
	Removed (%)	36.9	45.8	-	-	-
coif3	Retention (%)	62.3	-	-	-	-
	Removed (%)	37.7	-	-	-	-
sym4	Retention (%)	62.5	53.3	51.0	-	-
	Removed (%)	37.5	46.7	49.0	-	-
bior2.4	Retention (%)	60.4	51.3	-	-	-
	Removed (%)	39.6	48.7	-	-	-

Table 3. Energy Compaction with Threshold C(i) > = 0.1

Wavelet	Energy compaction	Decomposition Level				
		1	2	3	4	5
Haar	Retention (%)	58.8	49.3	46.9	46.3	46.1
	Removed (%)	41.2	50.7	53.1	53.7	53.9
db5	Retention (%)	60.9	50.9	-	-	-
	Removed (%)	39.1	49.1	-	-	-
coif3	Retention (%)	59.8	-	-	-	-
	Removed (%)	40.2	-	-	-	-
sym4	Retention (%)	60.0	50.7	48.4	-	-
	Removed (%)	40.0	49.3	51.6	-	-
bior2.4	Retention (%)	59.1	49.8	-	-	-
	Removed (%)	40.9	50.2	-	-	-

For removing the coefficients $C(i) < 0.1$ as well as for $C(i) < 0.001$, Haar wavelet has the highest percentage of coefficients removed, for which significant coefficients (retention) are less than other wavelets (db5, coif3, sym4, dan bior2.4).

According the previous mentioned reason, based on recognition level as well as its energy compaction, Haar wavelet has good performance for extracting the iris feature when compared with other wavelets.

5. CONCLUSIONS

From results of tests and analysis, some conclusions can be drawn are:

1. The higher level of decomposition, the recognition level tends to be better
2. Applying Haar wavelet produces the highest recognition rate in all allowed decomposition level (from level 1 up to 5)
3. For removing the coefficients $C(i) < 0.1$ as well as for $C(i) < 0.001$, Haar wavelet has the highest percentage of coefficients removed, for which significant coefficients

(retention) are less than other wavelets (db5, coif3, sym4, dan bior2.4).

4. Based on recognition level as well as its energy compaction, Haar wavelet has good performance for extracting the iris feature when compared to other wavelets.

6. REFERENCES

[1] CASIA, *Iris Image Database*, www.sinobiometrics.com.

[2] Daugman, J. "How Iris Recognition Works", IEEE Transaction on Circuits and system for Video Technology, vol 14, no.1, January 2004.

[3] Daugman, J. "High Confidence Visual Recognition of Persons by a Test of Statistical Independence", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, No.11, pp. 1148-1161, 1993.

[4] Daugman, J. "*Iridology",* http://www.cl.cam.ac.uk/ users/jgd1000/, Februari, 2009.

[5] Drygajlo, A., "*Biometrics Lecture 7 Part 1-2008",* Speech Processing and Biometrics Group Signal Processing Institute Ecole Polytechnique Fédérale de Lausanne (EPFL), 2008.

[6] Du, Y., R. Ives, B. Bonney, and D. Etter, "*Analysis of Partial Iris Recognition"*, U.S. Naval Academy, 2004

[7] Haralick, R.M., K. Shanmugam, and I. Dinstein, "*Textural Feature for Image Classification",* IEEE Transaction on System, Man and Cybernetics Vol smc-3, November, 1973.

[8] Kusuma, A.A., "Pengenalan Iris Mata Menggunakan Pencirian Matriks Ko-Okurensi Aras Keabuan", Skripsi S-1, Universitas Diponegoro, Semarang, 2009.

[9] Listyaningrum, R., *Analisis Tekstur Menggunakan Metode Transformasi Wavelet*, Skripsi S-1, Universitas Diponegoro, Semarang, 2006.

[10] Ma L. and T. Tan, *Personal Identification Based on iris Textural Analysis*", IEEE Transaction on Pattern Analysis and machine intelligence, December 2003.

[11] Masek, L, *Recognition of Human Iris Pattern for Biometric Identification.* The University of Western Australia, 2003.

[12] Misiti, M., Y. Misiti, G. Oppenheim, and J.M. Poggi, *Wavelet Toolbox For Use with Matlab*, User's Guide Version 3, The Math Works, Natick, MA, 2004.

[13] Moreno, R.P. and A. Gonzaga, *Features Vector For Personal Identification Based On Iris Texture*. Departamento de Engenharia Elétrica - EESC – USP, Februari, 2009.

[14] Munir, R., *Pengolahan Citra Digital dengan Pendekatan Algoritmik*, Informatika, Bandung, 2004.

[15] Peng, T., *"Circular Hough Based on Gradient",* Departement of Mechanical Engineering University of Maryland, College Park, Maryland, USA, Maret, 2007.

[16] Putriningsih, D., "*Identifikasi Kelebihan Kolesterol Berdasarkan Pengamatan Citra Iris Mata", Skripsi S-1*, Universitas Diponegoro, 2007.

[17] Rahmawati, I., "*Pemampatan Citra Digital Dengan Wavelet Paket*", Skripsi S-1, Universitas Diponegoro, 2007.

[18] Wijayanto, W.S., *Identifikasi Iris Mata dengan Tapis Gabor Wavelet dan Jaringan Syaraf Tiruan Learning Vector Quantization (LVQ)*, Student of Diponegoro University, Semarang, 2005.

[19] ---, *Image Processing Toolbox, for User's with MATLAB*, User's Guide Version 3, The Mathwork Inc, 2001.

[20] Merry, RJE., M. Steinbuch, and M.J.G. van de Molengraft, *Wavelet Theory and Applications, a Literature Study*, Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, Eindhoven, 2005.

[21] Lees, K. and J. Winkler, *Image Compression Using Wavelets*, http://www.dcs.shef.ac.uk/intranet/teaching/ projects/ archive/ug2002/pdf/u9kvl.pdf, 11 Januari 2012.

Colors Reduction in Computer Vision

Marian S. Stachowicz
University of Minnesota, USA
Warsaw School of Informatics, Poland
1 - 218 - 726 6531
mstachow@d.umn.edu

ABSTRACT
Color provides a wealth of information about an image. Color reduction and color feature extraction can be used to compress this information into a manageable image characteristic. This upgraded image-processing algorithm can be used to quickly describe images while providing enough information to allow image identification and classification. This paper investigates two possible applications of color recognition: stamp identification and landscape classification. The results from these examples show that color reduction can provide a useful tool for image understanding. In addition, the low processor cost of this system allows it to be easily incorporated into larger systems.

Keywords
Color recognition, image identification, image classification

1. INTRODUCTION
Color is an important feature of images that has been largely ignored until recent times. Most research has been focused around finding the spatial content in images such as edges and shapes. These features are then used to identify the images. This type of image processing is generally slow and complicated [4], [5], [3]. The idea presented in this paper is the use of an image's color content to describe the image. The process of compressing the color information into a manageable feature is useful and fast. This paper presents ideas regarding how a basic color feature can be used to represent an image in two applications: stamp identification and classification of nature scenes.

1.1 Color and computer images

Throughout history different color systems have been developed to accomplish various tasks. There are three basic color systems that are most commonly seen.These systems are RGB, HSB, and CMYK. The CMYK color system uses the four colors: cyan, magenta, yellow, and black. This system is most commonly seen in color printing. The HSB system represents colors as a combination of hue, saturation, and brightness. This system is more in tune with the way humans perceive color; therefore, it is useful for describing color to people. The RGB color system uses the colors: red, green, and blue. This system is the most common system, because it is used in almost all man made devices, such as: computer monitors, television sets, and slides.

Since computer monitors use RGB color, it is common practice to store computer images in terms of RGB [2]. In this paper all images are described in terms of an RGB color matrix. The RGB system was found to be the easiest to use because the incoming images were already described in terms of RGB, and converting to other color systems can be costly [2].

Images in computers are generally represented by a matrix of RGB color elements. These elements, also called pixels, take on the form {r,g,b}where r, g, and b represent the levels of red, green, and blue in the pixel. The values for red, green, and blue can have different ranges depending on the type of image. The true color or 24-bit image is commonly used in computers. In a true color image the r, g, and b values range from 0 to 255. Eight bits are needed to specify a value between 0 and 255, and there are three components to an RGB pixel. Therefore, 24 bits are needed to hold the information for each pixel, and that is why the true color image is also referred to as a 24-bit image (Sun & He, 1998). A 24-bit image can contain 2^{24} or 16,777,216 possible colors.

The pixel is the basic element of an image. Each pixel can be described by its location in the x-y plane and its color. The number of pixels to a side is a common way to describe the size of an image. For example, the size of an image could be written as 100 x 150, meaning that the image is 100 pixels wide and 150 pixels high. The pixels can also be used to describe the resolution of an image. Dots per inch (DPI) are a measure determined by the number of pixels per inch. For example, if an image has a resolution of 100 DPI and it is 1 inch square then its dimensions could alternatively be written in pixel notation as 100 x 100 [2].

1.2 The color recognition method
The color recognition method is an efficient way to describe color images. The basic idea of the color recognition method is to compress the color information in an image into a small feature space that still contains enough information to separate one image from another. This method was developed by Dr. Marian S. Stachowicz and Sony Zhan [10]. The first step in this system is to color reduce in the image. The colors contained in the image are reduced from the possible 16,777,216 to eight: red, green, blue, cyan, magenta, yellow, white, and black. The eight-color reduction is accomplished with equation (1).

r, g, b = {0 if r, g, b < 128, otherwise 255 (1)

After the reduction has been completed, the next step is feature extraction. Feature extraction is the process by which an n-

dimensional feature space is reduced to an r-dimensional feature space, where typically r<<n. The key to doing this reduction is that the information available in the original n-dimensional feature space must be preserved or enhanced in the new r-dimensional feature space. In this case the original feature space is the color in the reduced image, and the new feature space is an eight-member vector. The vector is of the form{r, g, b, c, m, y, w, bk}, where the members of this vector are the ratios of the eight colors in the reduced image. Once the color-ratio vector is created, it can be used to represent the image in a database. In this paper, the databases are lists with elements of the form: {name, feature vector}. When a new image to be identified or classified is input into the system, a feature vector is created for the new image. The new feature vector is compared to a database and the name of the closest matching entry in the database is returned [7], [8], [9]. The distance function (2) is used to determine which entry in the database is most similar to the new image.

$$D=\mathrm{Sqrt}[(r_1-r_2)^2+(g_1-g_2)^2+(b_1-b_2)^2+(c_1-c_2)^2+(m_1-m_2)^2+(y_1-y_2)^2+(w_1-w_2)^2+(bk_1-bk_2)^2] \quad (2)$$

2. APPLICATIONS OF COLOR REDUCTION

A human being can discern between the flags, stamps, and landscapes with relative ease, but to a computer the problem of identifying these images is complicated. Images are identifiable by color and content. The content of an image is difficult to describe to a computer. Calculations used to determine image content such as edge detection and segmentation are complicated and time consuming [4]. The color information however, can be passed into a computer system quickly. This section will outline the use of the color recognition method on two different applications.

2.1 Stamp Identification

The stamp identification system was developed to locate and identify stamps on white envelopes. Alternatively, this system could be used to determine the amount of postage on an envelope. The stamp identification system takes the use of color further by using color information to assist in locating and extracting the stamps from the background envelope. For this example, the database contained 104 stamp images obtained from the Internet. The stamp identification was found to be up to 97% on test images, which consisted of two to four stamps on a white background. The graph on the left in Figure 1 shows the feature vectors for the stamps: aquarium_fish1,fruit_berries2, and Wisconsin. This graph is an example of how the database could be visualized. The three graphs all appear to be similar, yet enough difference exists between them to allow correct identification of the stamp images. It can be seen that the graph on the right side of Figure 1 corresponds to the image aquarium_fish1 in the graph to the right. This system is less prone to the problems found in the flag example presented in [7], because the stamps contain a greater diversity of color information. The main limitation to stamp location system is that the stamp image must be correctly located on the envelope.

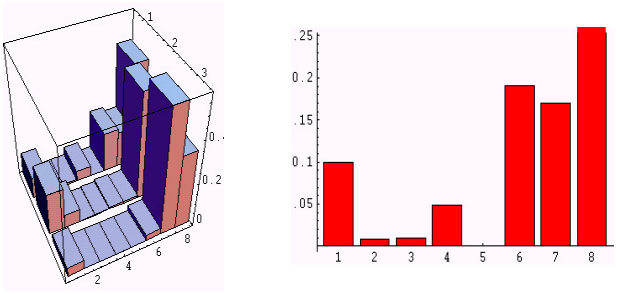


Figure 1. An example of a small database and an individual vector

2.2 Landscape classification

The process of image classification is slightly different from image identification. In image classification the images are divided into groups, which are referred to as classes. The images in each class are related to each other by some criteria. The database contains one feature vector for each class instead of each image. In an image classification system the input images need not be images that are in the database. The goal of the system is to put the new image into the correct class. For this example, three classes are used: Blue Sky, Green Plant, and Rock Mountain [10]. Two separate databases were created for the system using seven images for each class. The first database used feature vectors determined using K-Means clustering, and the feature vectors for the second database were found with Fuzzy C-Means clustering [11]. When an unknown image is input into both systems it is correctly classified [10].

3. CONCLUSION

The example listed in section 2 show that the technique of color recognition is useful in a number of applications. These examples also enforce the idea that color can be used effectively to represent an image without any spatial information. In addition, the examples demonstrate the low processor cost of the calculations performed on the input images. The color recognition system does have limitations. The system performs best on small databases of images that have a diverse color palette. Also, the performance of the system is decreased when the input images are missing parts, or have extra parts. The efficiency with which the color reduction method can describe images, makes it an effective tool on its own or coupled with a more robust computer vision system. The upgraded algorithm significantly increased speed of the color processing in the digital images.

4. ACKNOWLEDGMENTS

All graphs in this paper were created with Mathematica 8.0 (Wolfram, 2011). We would also like to thank Sony Zhan for the research he conducted on his Master's Thesis and David Olsen for his tutorial about image processing using Mathematica software.

5. REFERENCES

[1].Androutsos, D., Plataniotis, K. N., and Venetsanopoulos, A. N. (1999), A novel vector-based approach to color image retrieval using a vector angular-based distance measure, Comp.Vision and Image Under., Vol. 75, pp. 46-58.

[2]. Fortner, B. and Meyer, T. E. (1997) Number by Colors, Springer-Verlag, New York.

[3]. Haering, N. and Lobo, N. V. (1999), Features and classification methods to locate deciduous trees in images, Comp. Vision and Image Under., Vol. 75, pp. 133-149..

[4].Jiang, X. and Bunke, H. (1999), Edge detection in range images based on scan line approximation", Comp. Vision and Image Under. Vol 73, pp. 183-199.

[5]. Paschos, G. and Valavanis, K. P. (1999), A color texture based visual monitoring system for automated surveillance, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 29, pp. 298-307.

[6]. Sun, W. and He, Y. (1999), Spatial-chromatic clustering for color image compression, Proceedings of 1998 IEEE World Congress on Computation Intelligence, FUZZ-IEEE, pp. 1601-1604,

[7] Stachowicz, M.S. and Lemke, D. (2000),

Color Recognition, Proceedings of ITI 2000, pp. 133-141.

[8]. Stachowicz, M.S. and Lemke, D. (2000), Recognition of U.S. Postage Stamps, Proceeding of the NCUR 2000, pp. 1801-1809,

[9]. Wolfram, S (2011). Mathematica Book, Addison-Wesley Publish Company Inc., Redwood City California,

[10]. Zhan, Z. (1998), Fuzzy Mathematical Approach to Color Recognition, Master Thesis, University of Minnesota.

[11] Stachowicz, M.S. and Beall, L., (2003), Fuzzy Logic Package for *Mathematica*, Wolfram Research, Inc.

KSVD - Gradient Descent Method For Compressive Sensing Optimization

Endra

Department of Computer Engineering
Bina Nusantara University
Jl. KH. Syahdan No.9, Kemanggisan,
Jakarta, 11480, Indonesia
6221-5345830

endraoey@binus.edu

ABSTRACT

Compressive sensing is the recent technique of data acquisition where perfect reconstruction of signal can be made from far fewer samples or measurement than traditional Shannon-Nyquist sampling theorem. Compressive sensing exploits the sparsity of signal that is signal can be representated with a few non-zero coefficients by using a suitable basis or sparse dictionary. Compressive sensing also requires a small mutual coherence between measurement matrix and the basis or sparse dictionary in order to achieve a good reconstruction performance from a small number of measurement. Random matrix is usually used as compressive measurement because it has a small enough mutual coherence with many basis like Fourier and Wavelet. However, random matrix still can be optimized to achieve a smaller mutual coherence in order to improve the compressive sensing performance. In this paper, a joint optimization of measurement matrix and sparse dictionary was proposed to minimize the mutual coherence beetwen sparse dictionary and measurement matrix. We combined KSVD and Gradient-Descent methods to perform that joint optimization. The optimized measurement matrix and sparse dictionary from the joint optimization was used for image encoding to provide a compressive measurement. The simulation results showed that the joint optimization improved the quality of reconstructed image from compressive sensing compared to the unoptimized one and to the measurement matrix optimization only.

Keywords

Compressive Sensing; Sparse Dictionary; Measurement matrix; KSVD; Gradient Descent.

1. INTRODUCTION

Compressive/Compressed Sensing, the term that was coined by Donoho in [1], is a novel signal acquisition method that offers joint sensing and compression. Hence, compressive sensing just need a small number of measurement to reconstruct the signal rather than conventional method that is Shannon-Nyquist sampling theorem [2,3]. Compressive Sensing becomes one of the recent interesting fields in signal and image processing communities after the important works of Donoho, Candès, Romberg and Tao in [1,4-6] that has initiated the research of compressive sensing. It has been utilized in many different applications from medical [7-8], wireless sensor network and networked data [9-10], radar imaging [11-12], encryption [13-14], imaging [15-16], remote sensing [17-18] and inference problems such as detection, classification/estimation and filtering [19]. Compressive sensing is performed through a linear projection of the signal into a measurement matrix to yield a compressed measurement so the process of measurement and compression is taken simultaneously. This is different from the traditional compression where the signal is sampled by following the Shannon-Nyquist sampling theorem and then compress it by using transform coding as in JPEG-2000 [20]. Compressive sensing utilizes the sparsity of natural signal that is signal can be represented with a few coefficients by using the appropriate basis such as Fourier, Wavelet, Curvelet or a dictionary. The basis or dictionary must be incoherent with the measurement matrix so that the signal can be reconstructed from a few number of measurements [21]. Random matrix with high probability is incoherent with many basis [6], therefore it is often used as sensing matrix in the application that we have mentioned earlier. In [22] random matrix was used as projection matrix for color image compressed measurement. However, as was showed in [23] and [24] the random matrix can be optimized by minimizing the coherence with the basis or dictionary that is used to represent the signal and result the better signal reconstruction. The optimizing methods in [23-24] also was applied in [25] and [26] to optimize the measurement matrix that result the better compressive sensing performance than the unoptimized one. Optimization method in [24] is better compared to the method in [23], however further improvement can be done by using learned dictionary. In this paper, method in [24] and KSVD method [27] were used for dictionary learning and measurement matrix optimization simultaneously to yield a joint measurement matrix dictionary optimization. The results in this paper showed that the joint optimization provide the better compressive sensing performance than optimize the measurement matrix only.

2. COMPRESSIVE SENSING THEORY

Sparsity and incoherence play an important role in compressive sensing where they determine its performance [21]. A signal is said sparse if it has only a small number of nonzero components compared to its total length by using a proper basis or a dictionary. Compressive sensing framework mainly consists of two crucial parts; encoding (measurement/sensing) and decoding (reconstruction). Our focus in this paper is on the first part where we found that by joint optimizing the sensing process and sparse dictionary will more improve the reconstruction performance rather than only optimizing measurement matrix.

2.1 Sparsity and Incoherence

Consider a signal $x \in \Re^N$ and $\Psi = \lfloor \psi_1 ... \psi_K \rfloor$ is $\in \Re^{N \times K}$, where x can be written as :

$$x = \Psi\theta \tag{1}$$

where $\theta = [\theta_1 ... \theta_N]^T$ are coefficients vector that represent x in Ψ. If Ψ is orthonormal and $K = N$, it is called a basis and if $K > N$ it is called over-complete dictionary and usually only called dictionary. Signal x is said *S-sparse* if θ only has S non-zero coefficients and the others are zero. It is said compressible if the others coefficients has insignificant value compare to the S-significant coefficients. Sparse representation is used as principal method in data compression like as *JPEG-2000* [20]. Sensing or measurement matrix $\Phi = [\phi_1 ... \phi_m]^T \in \Re^{m \times N}$ is used to encode the signal x as a linear projection of it into Φ :

$$y = \Phi x = \Phi\Psi\theta \tag{2}$$

therefore y is a compressive measurement of x with length $m << N$. The coherency of Φ and Ψ is defined by :

$$\mu(\Phi, \Psi) = \sqrt{N} . \max_{\substack{1 \le i \le m, \\ 1 \le j \le N}} \left| \langle \phi_i, \psi_j \rangle \right| \tag{3}$$

where $\mu(\Phi, \Psi) \in \left[1, \sqrt{N}\right]$, assume Φ and Ψ are normalized. If the number of measurement fulfill :

$$m \ge C . \mu^2(\Phi, \psi) . S . \log(N) \tag{4}$$

where C is a positive constant and x is *S-sparse* , with overwhelming probability, x can be reconstructed from y exactly. If the support of θ (number of nonzero coefficients) fulfill :

$$\left\| \hat{\theta} \right\| < \frac{1}{2} \left(1 + \frac{1}{(\mu(\Phi, \Psi))} \right), \tag{5}$$

so again x can be reconstructed from y exactly. In (4) and (5) show that CS requires a small value of $\mu(\Phi, \Psi)$ in order to just need a few number of measurements and the sparsity condition is not too tight.

2.2 Measurement Matrix Optimization

The multiplication of sensing matrix and dictionary, $\Phi\Psi := D$ is called equivalent dictionary. *Elad* proposed in [23] the *mutual coherence* parameter, $\mu(D)$ that is defined by :

$$\mu(D) := \max_{i \ne j, 1 \le i, j \le K} \left\{ d_i^T d \right\} \tag{6}$$

that related to the coherency of Φ and Ψ, assume the column of D has a unit 2-norm . The sensing matrix can be optimized by minimizing $\mu(D)$, it is done by making the Gram matrix of D , $G = D^T D$ as close as identity matrix I :

$$D = \arg\min_D \left\| D^T D - I \right\|_F^2 \tag{7}$$

Vahid Abolghasemi et al. in [24] proposed to minimize the mutual coherence of equivalent dictionary by using Gradient Descent method. First by defining the corresponding error as :

$$E = \left\| D^T D - I \right\|_F^2 = Tr\left\{ \left(D^T D - I \right) \left(D^T D - I \right)^T \right\} \tag{8}$$

then compute the gradient of E with respect to elements of D that is $d_{i,j}$:

$$\nabla E = \frac{\partial E}{\partial d_{ij}} = 4D\left(D^T D - I \right) \tag{9}$$

using (9), the solution of (7) can be described as an iterative process to update D by using :

$$D_{(i+1)} = D_{(i)} - \eta D_{(i)} \left(D^T{}_{(i)} D_{(i)} - I \right) \tag{10}$$

where η is the step size. The updated sensing matrix can be achieved by using $\Phi = D\Psi^{-1}$ and before run to the next iteration step, D must be normalized again so that the column has a unit 2-norm. After a number of iterations the solution of (7) can be achieved and we get the optimized sensing matrix.

2.3 Gradient Descent-KSVD Method

Following the idea of joint optimization method in [28], the KSVD algorithm in [27] for dictionary learning and Gradient Descent method in [25] for measurement matrix optimization was combined to solve the following optimization problem :

$$\min_{\Psi, \Phi, \Theta} \left\{ \alpha \left\| X - \Psi\Theta \right\|_F^2 + \left\| Y - \Phi\Psi\Theta \right\|_F^2 \right\} \; s.t. \; \forall i, \left\| \theta_i \right\|_{\ell_0} \le S \tag{11}$$

where $X \in \Re^{N \times P}$ is a set of P training patch, where $0 \le \alpha \le 1$ is used to control the weigthed error of term $\left\| X - \Psi\Theta \right\|_F^2$ and Y is the projection of X into Φ

$$Y = \Phi X \tag{12}$$

where $Y = [y_1 ... y_P] \in \Re^{m \times P}$ with $m << n$. α is used in (11) to compensate the large reconstruction error of $\left\| X - \Psi\Theta \right\|_F^2$ and also to provide a more important role on projection error term $\left\| Y - \Phi\Psi\Theta \right\|_F^2$ that is actually happened in the stage of reconstruction. Eq. (11) can be written in the form :

$$\min_{\Psi,\Phi,\Theta}\left\|\begin{pmatrix}\alpha X\\ Y\end{pmatrix}-\begin{pmatrix}\alpha I\\ \Phi\end{pmatrix}\Psi\Theta\right\|_F^2 \quad s.t.\ \forall i,\|\theta_i\|_{\ell_0}\le S \qquad (13)$$

by defining $Z:=\begin{pmatrix}\alpha X\\ Y\end{pmatrix}$ and $W:=\begin{pmatrix}\alpha I\\ \Phi\end{pmatrix}$, (12) can be expressed as

$$\min_{\Psi,\Phi,\Theta}\left\|Z-D_{eq}\Theta\right\|_F^2 \quad s.t.\ \forall i,\|\theta_i\|_{\ell_0}\le S \qquad (14)$$

where $D_{eq}:=W\Psi=\left[d_1^{eq}...d_K^{eq}\right]$. The solution of (14) was conducted by combining Gradient Descent optimization of measurement matrix and KSVD. Initially, a certain of $\hat{\Psi}$ and $\hat{\Phi}$ are used, measurement matrix $\hat{\Phi}$ is optimized by using Gradient Descent method. *OMP* (*Orthogonal Matching Pursuit*) [29] is used to obtain $\hat{\Theta}$ from (13). Next, $\hat{D}_{eq}$ is found by using KSVD method and by using the optimized $\hat{\Phi}$ we can get the learned dictionary $\hat{\Psi}$ from $\hat{D}_{eq}$. The learned dictionary $\hat{\Psi}$ is used in the next iteration to optimize the measurement matrix $\hat{\Phi}$ and KSVD method is used again to get the learned dictionary $\hat{\Psi}$, the process is repeated for a number of iteration or until the stopping criteria is attained.

3. SIMULATION METHOD

We used the three different scenarios of compressive sensing :

1. *KSVD-Random* : Random matrix is used as measurement matrix Φ and KSVD algorithm is used to get the learned dictionary Ψ.
2. *Uncoupled KSVD-Gradient Descent* : Random matrix Φ is optimized by using Gradient Descent method and KSVD algorithm is used to get the learned dictionary Ψ, both of them were performed separately.
3. *Joint KSVD-Gradient Descent* : Joint Measurement Matrix-Dictionary Optimization was performed by using KSVD-Gradient Descent as was described in the previous section where initially random matrix is used.

We used 30 training-images (available in [30]) as shown in Fig. 1. From each image is taken randomly two hundred 8x8 patches so we have 6000 patches that are used as $X\in\Re^{N\times P}$ where $N=64$ and $P=6000$. This training patch X was used in Gradient Descent optimization and KSVD method to obtain the optimized measurement matrix Φ and the learned dictionary Ψ.

Figure 1. The 30 images that are used as training patches.

For all those scenarios, we used *Overcomplete Discrete Cosine Transform* (Overcomplete *DCT*) as initial dictionary Ψ with $K=4N=256$. We used a test image to perform those scenarios where to reduce the computational complexity in compressive sensing reconstruction, the image was divided into 8x8 non-overlapping blocks. The next step is each block convert to 1-D signal, so $N=64$, compressive measurement was performed on those blocks by using Φ from one of the three scenarios above to yield $Y\in\Re^{m\times P}$. The random matrix $\Phi\in\Re^{m\times 64}$ was generated with i.i.d Gaussian elements and normalized by $\sqrt{N}=8$ where m is number of measurement that was varied from 10 to 25. We used $\eta=0.01$ for Gradient Descent and $\alpha=\frac{1}{32}$ for joint optimization. The reconstruction of each block was performed by using *OMP* and for comparison also used *Iteratively Reweighted Least Squares (IRLS) - ℓ_p - minimization* [31] with $p=0.8$ to get Θ from Y and reconstructed blocks are obtained by using $X=\Psi\Theta$. Finally we deblocked the whole reconstructed blocks to get the reconstructed image I. We used the Peak Signal-to-Noise Ratio (PSNR) in Decibel *(dB)* to measure the reconstruction performance, it is defined in (15) , where W and H are width and height of image.

$$PSNR=10\log_{10}\left(\frac{W\times H\times 1^2}{\sum_{w=1}^{W}\sum_{h=1}^{H}\left(I_{w,h}^{reconstructed}-I_{w,h}^{test}\right)^2}\right)dB \qquad (15)$$

4. RESULTS & DISCUSSION

We used $W\times H=481\times 321$ "*Sphinx*" as a test image as shown in Fig. 2.

Figure 2. The "*Sphinx*" that was used as a test image to perform the three different scenarios of compressive sensing.

Fig. 3 and Fig. 4 show the PSNR of reconstructed image from compressive sensing by using *OMP* and *IRLS* - ℓ_p - *minimization* as a function of measurement numbers m = 10 to 25 for the three different scenarios above. For reconstruction by using *OMP*, it showed that by optimizing the measurement matrix can improve the PSNR of reconstructed image where *Joint KSVD-Gradient Descent* outperformed the other scenarios. The PSNR increment of *Joint KSVD-Gradient Descent* is up to 145 % at m = 10 compared to the *KSVD-Random* and up to 15 % at m = 15 and 16 compared to the *Uncoupled KSVD-Gradient Descent*.

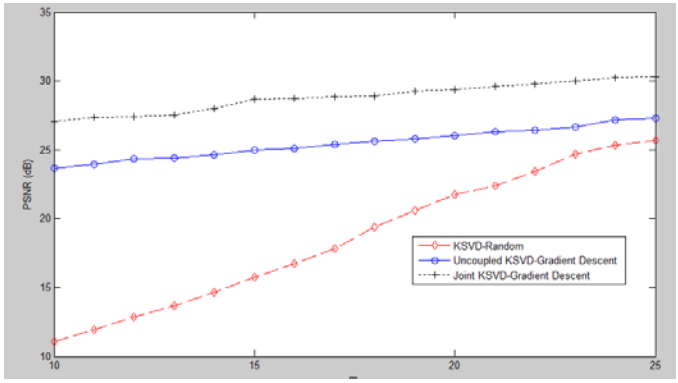


Figure 3. The PSNR comparison of reconstructed image from compressive sensing by using *OMP* for : *KSVD-Random, Uncoupled KSVD-Gradient Descent* and *Joint KSVD-Gradient Descent.*

Reconstruction by using *IRLS* - ℓ_p - *minimization* provided higher PSNR than *OMP* but with longer computation time. The result is consistent where *Joint KSVD-Gradient Descent* outperformed the other scenarios. Fig. 4 also showed that *Uncoupled KSVD-Gradient Descent* method improved the PSNR of reconstructed image than *KSVD-Random* but only up to m = 17, after that both of them provided almost the same PSNR. The PSNR increment of *Joint KSVD-Gradient Descent* is up to 69 % at m = 10 compared to the *KSVD-Random* and up to 12 % at m = 10, 11, 12 and 13 compared to the *Uncoupled KSVD-Gradient Descent*.

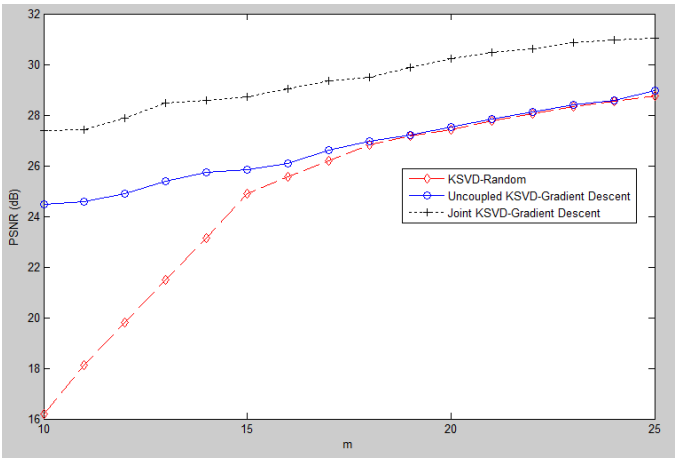
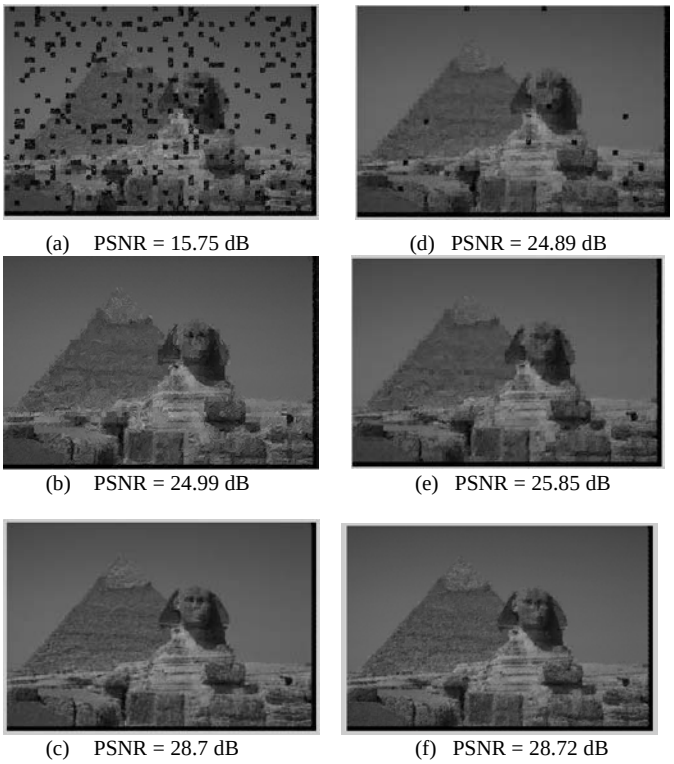


Figure 4. The PSNR comparison of reconstructed image from compressive sensing by using (IRLS) - ℓ_p - minimization for : KSVD-Random, Uncoupled KSVD-Gradient Descent and Joint KSVD-Gradient Descent.



(a) PSNR = 15.75 dB (d) PSNR = 24.89 dB

(b) PSNR = 24.99 dB (e) PSNR = 25.85 dB

(c) PSNR = 28.7 dB (f) PSNR = 28.72 dB

Figure 5. The comparison of reconstructed image for m = 15 by using *OMP* (left column) and *IRLS* - ℓ_p - minimization (right column) where : (a) & (d) *KSVD-Random*, (b) & (e) *Uncoupled KSVD- Gradient Descent* , (c) & (f) *Joint KSVD- Gradient Descent.*

Fig. 5 shows the reconstructed image from compressive sensing for measurement numbers m = 15 where the *Ratio Measurement Numbers* $(RMN) = \frac{m}{N} x100\% = 23.44\%$ for the three scenarios above. It showed that by optimizing the measurement matrix, the

reconstructed image is also improved visually where for *Joint KSVD-Gradient Descent* outperformed than the others.

5. CONCLUSION

In this paper, we proposed joint sparse dictionary - measurement matrix optimization by using KSVD - Gradient Descent method to improve the reconstruction image performance from compressive sensing. From the results and discussions, it showed that by optimizing measurement matrix and dictionary learning simultaneously provided the improvement of the image reconstruction from compressive sensing in PSNR and visually. Further improvement can be attempted in future work by optimizing measurement matrix and dictionary learning simultaneously based on block-sparse representations.

6. REFERENCES

[1] David L. Donoho, Apr. 2006, *Compressed Sensing*, IEEE Transactions on Information Theory., vol. 52, no. 4, pp. 1289-1306.

[2] Michael Unser, Apr. 2000, *Sampling—50 Years After Shannon*, Proceedings of the IEEE., vol. 88, no. 4, pp. 569-587.

[3] Candes, E.J., and Wakin, M.B., March. 2008, *An Introduction to Compressive Sampling*, IEEE Signal Processing Magazine., pp. 21-30.

[4] Emmanuel J. Candès, Justin Romberg, and Terence Tao, Feb. 2006, *Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information*, IEEE Transactions on Information Theory., vol. 52, no. 2, pp. 489-509.

[5] E. Candès, J. Romberg, and T. Tao, Aug. 2006 , *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223.

[6] Emmanuel J. Candès and Terence Tao, Dec. 2006, *Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*, IEEE Transactions on Information Theory., vol. 52, no. 12, pp. 5406-5425.

[7] Jing Wu and Ye Li, Nov. 2009, *Low-complexity Video Compression for Capsule Endoscope Based on Compressed Sensing Theory*, in Proc. International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009., pp. 3727-3730.

[8] Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M., March. 2008, *Compressed Sensing MRI*, IEEE Signal Processing Magazine., pp. 72-82.

[9] Haupt, J., Bajwa, W.U., Rabbat, M., and Nowak, R., March. 2008, *Compressed Sensing for Networked Data*, IEEE Signal Processing Magazine., pp. 92-101.

[10] Peng Zhang, Chen Chen, and Minrun Liu, Nov. 2009, *The Application of Compressed Sensing in Wireless Sensor Network*, in Proc. International Conference on Wireless Communication & Signal Processing, WCSP 2009., pp. 1-5.

[11] Lei Yu, Yi Yang, Hong Sun, and Chu He, Oct. 2009, *Turbo-like Iterative Thresholding for SAR Image Recovery from Compressed Measurements*, in Proc.2nd Asian Pacific Conference on Synthetic Aperture Radar, APSAR 2009., pp. 664-667.

[12] Matthew A. Herman and Thomas Strohmer, Jun. 2009, *High-Resolution Radar via Compressed Sensing*, IEEE Transactions on Signal Processing., pp. 2275-2284.

[13] A Anil Kumar and Anamitra Makur, Jan. 2009, *Lossy Compression of Encrypted Image by Compressive Sensing Technique*, in Proc. IEEE Region 10 Conference TENCON 2009., pp. 1-5.

[14] Adem Orsdemir, H. Oktay Altun, Gaurav Sharma, and Mark F. Bocko, Nov. 2008, *On The Security and Robustness of Encryption Via Compressed Sensing*, in Proc. IEEE Military Communications Conference, MILCOM 2008., pp. 1-7.

[15] Justin Romberg, March. 2008, *Imaging via Compressive Sensing*, IEEE Signal Processing Magazine., pp. 14-20.

[16] Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J.N., Ting Sun, Kelly, K.F., and Baraniuk, R.G., March. 2008, *Single-Pixel Imaging via Compressive Sampling*, IEEE Signal Processing Magazine., pp. 83-91.

[17] Jianwei Ma, Oct. 2009, *A Single-Pixel Imaging System for Remote Sensing by Two-Step Iterative Curvelet Thresholding*, IEEE Geoscience and Remote Sensing Letters., vol. 6, no. 4, pp. 676-680.

[18] Jianwei Ma, Apr. 2009, *Single-Pixel Remote Sensing*, IEEE Geoscience and Remote Sensing Letters., vol. 6, no. 2, pp. 199-203.

[19] Mark A. Davenport, Petros T. Boufounos, Michael B. Wakin, and Richard G. Baraniuk, Apr. 2010, *Signal Processing With Compressive Measurements*, IEEE Journal of Selected Topics in Signal Processing., vol. 4, no. 2, pp. 445-460.

[20] D.S. Taubman and M.W. Marcellin, 2001, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*, Norwell, MA: Kluwer.

[21] E. Candès and J. Romberg, 2007, *Sparsity and incoherence in compressive sampling*, Inverse Prob., vol. 23, no. 3, pp. 969–985.

[22] Endra, Oct. 2010, *Color Image Reconstruction From Compressive Sensing Using Iteratively Reweighted Least Squares-* ℓ_p *– Minimization*, in Proc. Makassar International Conference on Electrical Engineering and Informatics, 2010.

[23] M. Elad, Dec. 2007, *Optimized projections for compressed sensing*, IEEE Transactions on Signal Processing, vol. 55, no. 12, pp. 5695–5702.

[24] Vahid Abolghasemi, Saideh Ferdowsi, Bahador Makkiabadi and Saeid Sanei, "On Optimization of The Measurement Matrix For Compressive Sensing," *in Proc. 18th European Signal Processing Conference, EUSIPCO 2010*, August 23-27, 2010.

[25] Endra, *Color Image Reconstruction from Compressive Sensing with Optimized Measurement Matrix*, Binus ICT National Conference, 15 – 16 Agustus 2011.

[26] Endra, *Comparison of Optimized Random Gaussian and Partial Random Fourier Measurement in Compressive Sensing*, 12th International Conference on Quality in Research (QiR), 4 – 7 July, 2011, University of Indonesia.

[27] M. Aharon, M. Elad, and A. Bruckstein, *The KSVD: An algorithm for designing overcomplete dictionaries for sparse representation*, IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[28] J. M. Duarte-Carvajalino and G. Sapiro, *Learning to sense sparse signals: simultaneous sensing matrix and sparsifying dictionary optimization*, IEEE Transactions on Image Processing, vol. 18, no. 7, pp. 1395–1408, July 2009.

[29] J. A. Tropp and A. C. Gilbert, 2007, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Transactions on Information Theory, vol. 53, no. 12, pp. 4655–4666.

[30] The Berkeley Segmentation Dataset and Benchmark, Available: http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/.

[31] Rick Chartrand and Wotao Yin, Apr. 2008, *Iteratively Reweighted Algorithms for Compressive Sensing*, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008., pp.3869-387.

Improved Speaker Identification with Gaussian Mixture Models (GMMs)

Smarajit Bose
Bayesian and Interdisciplinary Research Unit
Indian Statistical Institute
Kolkata, India
+91 33 25753412
smarajit@isical.ac.in

Amita Pal
Bayesian and Interdisciplinary Research Unit
Indian Statistical Institute
Kolkata, India
+91 33 25752817
pamita@isical.ac.in

Debapriya Sengupta
Applied Statistics Unit
Indian Statistical Institute
Kolkata, India
+91 33 25753438
dps@isical.ac.in

Gopal K. Basak
Statistics and Mathematics Unit
Indian Statistical Institute
Kolkata, India
+91 33 25753456
gkb@isical.ac.in

ABSTRACT
Speaker identification using the highly successful Gaussian Mixture Models (GMMs) based on Mel Frequency Cepstral Coefficients (MFCCs) as features, proposed by Reynolds and Rose (1995), has been universally acknowledged to be superior to many other existing approaches. The use of GMMs for modeling speaker identity is motivated by the interpretation that the Gaussian components represent some general speaker-dependent spectral shapes and the capability of mixtures to model arbitrary densities. In this work, we have demonstrated, with help of benchmark data sets like NTIMIT, how the well-known principal component transformation can be exploited to enhance the performance of the MFCC-GMM speaker recognition systems.

Keywords
Mel frequency cepstral coefficients, Gaussian mixture models, principal component transformation, classification accuracy, NTIMIT

1. INTRODUCTION
Automatic speaker identification/recognition (ASI/ASR) is the generic term applied to the automatic process of inferring the identity of a person from an utterance made by him, on the basis of speaker-specific information embedded in the corresponding speech signal. This technique has important practical applications, e.g., it can be used to verify the identity claimed by people accessing systems, that is, it enables access control of various services by voice. Other real-life activities where it is immediately applicable and useful include voice dialing, banking over a telephone network, telephone shopping, database access services, information and reservation services, voice mail, security control for confidential information, and remote access to computers. Another important application of speaker recognition technology is in forensics.

Speaker recognition, being essentially a pattern recognition problem, can be specified broadly in terms of the features used and the classification technique adopted. From experience gained over the past several years from research going on, it has been possible to identify certain features extracted from the complex speech signal, that carry a great deal of speaker-specific information. In conjunction with these features, researchers have also identified classifiers which perform admirably. Mel Frequency Cepstral Coefficients (MFCCs) and Linear Prediction Cepstral Coefficients (LPCCs) are the popularly used features, while Gaussian Mixture Models (GMMs), Hidden Markov Models (HMMs), Vector Quantization (VQ), Neural Networks are some of the more successful speaker models/classification tools. Any good review article on speaker recognition (for example, [1], [2]) contains details and citations about more than a few of these features and models. It is quite apparent that much of the research involves juggling various features and speaker models in different combinations to get new ASR methodologies.

Reynolds [3] proposed a speaker recognition system based on MFCCs as features and GMMs as speaker models and by implementing it on the benchmark data sets TIMIT and NTIMIT, demonstrated that it works almost flawlessly on clean speech (TIMIT) and quite well on noisy telephone speech NTIMIT). This approach is still one of the best available in the literature.

In this paper, we have established with the help of NTIMIT, how the classification accuracy of the basic MFCC-GMM speaker recognition system can be enhanced significantly by incorporating into the model something that Reynolds ignored totally by assuming the MFCCs to be independent, namely, the individual correlation structures of the feature sets for each speaker. This is achieved by the simple device of the Principal Component Transformation, which is applied to the covariance matrix of the MFCCs for each speaker. Due to differences in the correlation structures, these transformations are also different for different speakers. The GMMs are fitted on the transformed variables using the principal component transformations instead of the original MFCCs. For testing, to determine the likelihood values for a particular speaker model, the MFCCs of the test utterance are transformed by the principal component transformation of the

corresponding speaker. Implementation of the proposed approach on NTIMIT shows a marked improvement in classification accuracy relative to the basic MFCC-GMM approach.

The paper is organized as follows. MFCCs are introduced in the following section, while Gaussian Mixture Models (GMMs) are briefly described in Section 3, which also outlines how speaker recognition is carried out using MFCCs as features and GMMs as speaker models. The proposed approach is delineated in Section 4. Section 5 gives a brief description of the benchmark NTIMIT dataset and contains some results obtained by applying the proposed approach on it, which clearly establish its effectiveness.

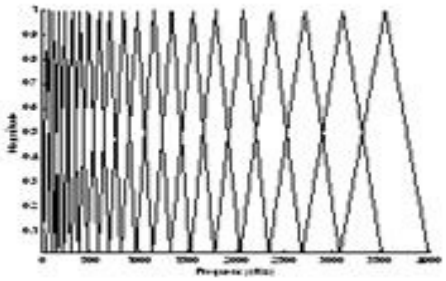


Figure 1. A Mel scale filter bank

2. MEL FREQUENCY CEPSTRAL COEFFICIENTS (MFCCs)

The Mel Frequency Cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log-energy spectrum on a nonlinear *mel* scale of frequency.

Mel-frequency cepstral coefficients (**MFCCs**) [3] are coefficients that collectively make up an MFC. Based on a bank of *K* filters, a set of *M* MFCCs are computed as

$$MFCC_i = \sum_{k=1}^{K} X_k \cos\left[i\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right], \quad i = 1,2,\mathrm{K}\ ,M,$$

being the log-energy output of the *k*-th filter.

For each speech frame, the energy X_k at time *n* for the *l*-th mel-scale filter is

$$E_{mel}(n,l) = \frac{1}{A_l}\sum_{k=L_l}^{U_l} |V_l(\omega_k)Y(n,\omega_k)|^2$$

where V_l is the response of the *l*-th mel-scale filter,

$$Y(n,\omega_k) = x[m]w[n-m]e^{-j\frac{2\pi}{k}m}$$

and

$$A_l = \sum_{k=L_l}^{U_l} |V_l(\omega_k)|^2 .$$

x[*m*], *w*[*m*] are respectively the speech signal and a window function at time point *m*.

Mel Scale Filter Banks

One way to simulating the spectrum is by using a filter bank, spaced uniformly on the mel scale (described below) (Figure 1). That filter bank has a triangular bandpass frequency response, and the spacing as well as the bandwidth is determined by a constant mel frequency interval.

The Mel Scale

Psychophysical studies show that human perception of the frequency contents of sounds for speech signals does not follow a linear scale.

For each tone with an actual frequency, *f*, measured in *hz*, a subjective pitch is measured on the so-called 'mel' scale.

The mel scale is a scale of pitches judged by listeners to be equal in distance from one another. The word *mel* comes from the word *melody* to indicate this**.** Th**is** scale is a linear frequency spacing below 1000 *hz* and a logarithmic spacing above 1000 *hz* **(Figure 2).**

A popular formula to convert *f* hertz into *m* mel is:

$$m = 2595\log(1+f/700)$$

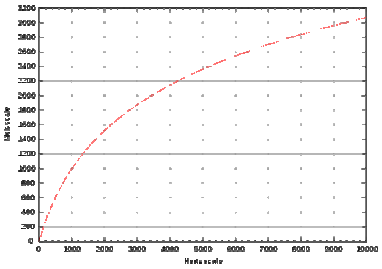


Figure 2. The mel scale

2.1 Computation of MFCCs

This involves the following steps:

1. Partitioning the speech signal into overlapping segments or frames
2. Taking the Fourier transform of signal from each frame.
3. Mapping the powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
4. Taking the logs of the powers at each of the mel frequencies.
5. Taking the discrete cosine transform of the list of mel log powers, as if it were a signal.

The MFCCs are the amplitudes of the resulting spectrum.

3. SPEAKER RECOGNITION WITH MFCC-BASED GMM SPEAKER MODELS

Gaussian mixture models (GMMs)

If $\boldsymbol{x}$ is a d-dimensional feature vector, then for a K -speaker problem, speaker i, i=1,2,…,K is modeled as a mixture of N component densities as follows:

$$p(\mathbf{x} \mid \boldsymbol{\lambda}_i) = \sum_{j=1}^{N} p_{ij}\, f_j(\mathbf{x} \mid \boldsymbol{\theta}_{ij}), \quad \sum_{j=1}^{N} p_{ij} = 1,$$

where, for the i-th speaker, p_{ij} is the prior probability for the j-th component of the mixture, $f_j(\mathbf{x} \mid \boldsymbol{\theta}_{ij})$ is the probability density of $\boldsymbol{x}$ in the j-th component, assumed to be Gaussian in this case. That is, for a GMM,

$$p(\boldsymbol{x} \mid \boldsymbol{\lambda}_i) = \sum_{j=1}^{N} p_{ij} \frac{1}{(2\pi)^{\frac{d}{2}} \left|\boldsymbol{\Sigma}_{ij}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{ij})' \boldsymbol{\Sigma}_{ij}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_{ij})},$$

where $\boldsymbol{\lambda}_i = \{p_{ij}, \boldsymbol{\theta}_{ij}, j = 1,2,\mathrm{K}\ ,N\}$ is the collection of unknown parameters for the i-th speaker. Thus for a GMM,

$$\boldsymbol{\theta}_{ij} = \{\boldsymbol{\mu}_{ij}, \quad \boldsymbol{\Sigma}_{ij}\}, \ i = 1,2,\mathrm{K}\ , K, \ j = 1,2,\mathrm{K}\ , N.$$

GMM models for all speakers are trained by the Expectation-Maximization algorithm [3]. An unknown speech sample is split into a number of overlapping segments, with MFCCs computed from each segment. Likelihood function for the unknown sample is computed, based on all MFCC vectors obtained from it. The unknown sample is classified by maximum likelihood.

4. NEW APPROACH

In this section we describe the proposed modification of the MFCC-GMM approach which has been effective in increasing the prediction accuracy significantly. In the usual approach, the MFCCs are assumed to be independent which was found to be very unrealistic. In fact some of them are often highly correlated. Further the correlation structures of these for different speaker models appear to be quite different.

Since incorporating all the correlation coefficients in the model would lead to a substantial increment in the computational complexity and the variance of the predictions, we came up with a novel idea. For each of the speakers, we consider the covariance matrix of the MFCC, and consider its Principal Component Transformation. Due to differences in the correlation structures, these transformations are also different for different speakers. The GMMs are fitted on the transformed variables using the principal component transformations instead of the original MFCCs. While classifying a test utterance, we need to calculate the likelihood scores for each of the speaker models. In this case for finding the likelihood values for a particular speaker model, we transform the MFCCs of the test utterance by the principal component transformation of the corresponding speaker.

As the results in the following section demonstrate, this approach brings about a marked improvement in the matching between the test utterance and the GMM of the actual speaker.

The Principal Component Transformation

This is a widely-used linear orthogonal transformation for converting a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called **principal components**. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Depending on the field of application, it is also called the Karhunen–Loève transform (KLT), the Hotelling transform, and so on.

5. RESULTS

The Benchmark Dataset NTIMIT

NTIMIT, together with TIMIT is an acoustic-phonetic speech corpus in English, belonging to the Linguistic Data Consortium (LDC) of the University of Pennsylvania. TIMIT consists of clean microphone recordings of 10 different read sentences (2 ***sa***, 3 ***si*** and 5 ***sx*** sentences, some of which have rich phonetic variability), uttered by 630 speakers (438 males and 192 females) from eight major dialect regions of the USA. It is characterized by 8-kHz bandwidth and lack of intersession variability, acoustic noise, and microphone variability or distortion. These features make TIMIT a benchmark of choice for researchers in several areas of speech processing.

NTIMIT, on the other hand, is the same speech from the TIMIT database played through a carbon-button telephone handset and recorded over local and long-distance telephone loops. This provides speech identical to TIMIT, except that it is degraded through carbon-button transduction and actual telephone line conditions. Performance differences between identical experiments on TIMIT and NTIMIT are therefore, expected to arise primarily from the degrading effects of telephone transmission.

Results with NTIMIT

The overall classification accuracy obtained by us with all 630 speakers in NTIMIT was 35% when 6 out of the 10 recordings per speakers were used for training and the remaining 4 were used for testing. The corresponding figure with 8 training utterances and 2 test utterances per speaker, is 42%.

Table 1. Results on all 630 NTIMIT Speakers with 8 training utterances and 2 test utterances

Experiment	Threshold for frame energy	No. of MFCCs	No. of Eigen Vectors	Minimum frequency	Maximum frequency	Percentage of reduction of frames (for test)	No. of correctly classified test utterances	Accuracy %
1	0.0	38	37	0	550	13.5	565	44.8
2	0.0	38	37	0	550	15	566	44.9
3	0.0	38	37	200	400	13.5	615	48.8
4	0.0	38	37	200	400	15	618	49.0
5	0.01	38	37	0	550	13.5	603	47.8
6	0.01	38	37	0	550	15	590	46.8
7	0.01	38	37	200	400	13.5	649	51.5
8	0.01	38	37	200	400	15	642	50.9
9	0.014	38	37	0	550	13.5	595	47.2
10	0.014	38	37	0	550	15	595	47.2
11	0.014	38	37	200	400	13.5	*659*	***52.3***
12	0.014	38	37	200	400	15	957	52.1
13	0.015	38	37	0	550	13.5	578	45.8
14	0.015	38	37	0	550	15	587	46.5
15	0.015	38	37	200	400	13.5	652	51.7
16	0.015	38	37	200	400	15	644	51.1

The number of MFCCs and filters were both taken to be 38. The competing classifiers were obtained by varying some parameters of the classifiers, whose values are given in the tables.

Tables 1 and 2 present results obtained on all 630 NTIMIT speakers by our approach with 8 and 6 training utterances respectively, on a number of competing MFCC-GMM classifiers. These were obtained by varying certain tuning parameters of the generic MFCC-GMM classifier; the values of the parameters are mentioned in the tables. The number of MFCCs and filters were both taken to be 38 for all these classifiers. The shaded rows in the tables identify the best results obtained by our approach. It is evident that there is a substantial improvement in classification accuracy, which is of the order of 7.2% with 6 training utterances and 10.3% with 8 training utterances.

Our investigations show that the computation due to these additional transformations seems to require only about 7.4% more time on an average than that required by the ordinary MFCC-GMM algorithm.

In this context, it should be noted that though Reynolds [3], one of the proponents of the MFCC-GMM approach, reported a recognition accuracy of 60.7% on NTIMIT, this performance has not been replicated by other people working in this area, not even the renowned speech processing group of IIT Madras, of which Prof. Hema Murthy is an important member. In a personal communication (by e-mail via a student) and meetings in which she was present, she informed us that with 32-component GMMs as speaker models, based on 6 training utterances and 4 test utterances, the best they could get was 36%. In fact, using a 64-component mixture with the same training-test data combination, her group has reported 40% accuracy, still nowhere near that reported by Reynolds. However, what is important that, if one could replicate the performance reported by Reynolds, then that too could be significantly improved upon by the application of the proposed approach.

Table 2. Results on all 630 NTIMIT Speakers with 6 training utterances and 4 test utterances

Experiment	Threshold for frame energy	No. of MFCCs	No. of Eigen Vectors	Minimum frequency	Maximum frequency	Percentage of reduction of frames (for test)	No. of correctly classified test utterances	Accuracy %
1	0.0	38	37	0	550	15	915	36.3
2	0.0	38	37	200	400	13.5	989	39.2
3	0.0	38	37	200	400	15	986	39.1
4	0.01	38	37	0	550	13.5	101	40.1
5	0.01	38	37	0	550	15	100	39.6
6	0.01	38	37	200	400	13.5	104	41.2
7	0.01	38	37	200	400	15	103	40.9
8	0.014	38	37	0	550	13.5	970	38.4
9	0.014	38	37	0	550	15	957	37.9
10	0.014	38	37	200	400	13.5	***106***	***42.2***
11	0.014	38	37	200	400	15	104	41.4
12	0.015	38	37	0	550	13.5	986	39.1
13	0.015	38	37	0	550	15	973	38.6
14	0.015	38	37	200	400	13.5	105	41.7
15	0.015	38	37	200	400	15	104	41.3
16	0.011	38	37	200	400	13.5	105	41.9

6. CONCLUSION

From the results presented in the previous section, it is quite evident that the modified approach using Principal Component Transformation can improve the performance of the Gaussian Mixture Model based on the Mel Frequency Cepstral Coefficients for speaker identification. This modification takes care of the inappropriateness of the usual independence assumption of the Cepstral Coefficients. This approach is particularly useful for rectifying the degradation of performance in noisy data.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the funding provided for this work by Department of Information Technology, Ministry of Communication and Information Technology, Government of India. They are also indebted to Prof. Hema Murthy of the Indian Institute of Technology Madras for her invaluable assistance in the form of useful discussions and for loan of software related to the MFCC-GMM speaker recognition system.

8. REFERENCES

[1] Furui, S. 1997 Recent advances in speaker recognition**.** Pattern Recognition Letters, Vol. 18, pp. 859-872.

[2] Campbell (Jr.) J. P. 1997 Speaker recognition: A tutorial. Proceedings of the IEEE, Vol. 85, No. 9, pp. 1437-1462.

[3] Reynolds, D.A. 1995. Large population speaker identification using clean and telephone speech. IEEE Signal Processing Letters, Vol. 2 , No. 3 46-48, DOI=10.1109/97.372

Java Characters Word Processing

Rudy Adipranata
Informatics Department
Petra Christian University
Surabaya, Indonesia
rudya@petra.ac.id

Gregorius Satia Budhi
Informatics Department
Petra Christian University
Surabaya, Indonesia
greg@petra.ac.id

Rudy Thedjakusuma
Informatics Department
Petra Christian University
Surabaya, Indonesia

ABSTRACT
Java writing has its own unique characteristics compare to Latin writing. Therefore, Java writing is more complicated to be learned and applied. Along with the development of technology, the typing method of Java characters could be helped by creating Java characters word processing software which is helpful to write Java characters more easily. Also, the software will be very helpful to preserve cultural values. This word processing has been tested and running well to produce the raw Java characters include basic, supplementary and '*pasangan*' characters.

Keywords
Word processing, Java characters

1. INTRODUCTION
Indonesia is a country which composed of many islands. Indonesia has a national language, Bahasa Indonesia. In addition to Bahasa Indonesia, there are many local languages in each region. The diversity of local languages is a cultural value that should be preserved for new generations.

Javanese language is one of the local languages that is widely used instead of Bahasa Indonesia. Javanese language has unique Java characters and different compared to Latin characters. Because of its uniqueness, Java characters are more complicated to learn and apply in the community. Along with the development of technology, the typing method of Java characters could be helped by creating Java characters word processing software which is helpful to write Java characters more easily. In this paper, we present the development of Java characters word processing that can produce basic, supplementary and '*pasangan*' characters.

2. WORD PROCESSING
Word processing is an application or computer software used for creating, editing, and printing documents. Word processing software can combine images, graphics and text [2]. In addition to these advantages, there are several other advantages, namely, insert text: text to include in any document section, delete text: to remove words, cut and paste: to remove (cut) some text from one place between the document and paste elsewhere, copy and paste: copy some text from one place and paste to another place, find: search the text in document, print: sends document to printer.

3. AUTOMATA
Automata are abstract machines that can recognize, accept, or generate a sentence in a particular language [2]. Automata is also a system consists of a finite number of states of the abstract machine that receives input and generate output in the form of a discrete issue (one at a time).The state is a condition that describe information about the last input. State is also regarded as the machine's memory. The input of the automata considered the language to be recognized by the machine. After receive input, automata engine will determine whether the input is acceptable or not acceptable. Automata in a simple circuit are shown at Figure 1.

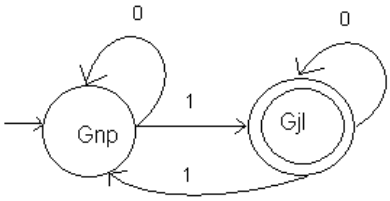


Figure 1. Simple Circuit Automata

Characteristics of automata: receiving input, generate output, has a buffer, able to make decisions in transforming inputs into outputs. Automata are also distinguished by the type of temporary memory it has, consists of three types: finite automata, pushdown automata and turing machine [2].

Finite automata don't have a temporary memory. Finite automata are a class of machines with the most limited abilities. Finite automata are used to aid the design of lexical analyzer, a text editor application, pattern recognition, fault tolerant systems, etc.

Pushdown automata (PDA) have LIFO (last in, first out) temporary memory mechanism. This engine is more powerful because of the presence of stack memory unit. Pushdown automata are used to identify the language that has context-free grammar, dictionary data, query, scripting, parsing, etc.

Turing machine has random access temporary memory mechanism. Turing machine can be used to identify the impossibility of writing a computer program. Turing machine is a mathematical model that is widely used for today's computers.

Some basic understanding of automata: the symbol is an abstract entity (similar to point in geometry). A letter or a number is an example of a symbol. String is a finite row symbols. For example, if a, b, and c are three symbols then abcb is a string constructed from those symbols. If w is a string then the string length is expressed as $|w|$ and defined for the number of symbols that make up the string. For example, if w = abcb then $|w| = 4$. Empty string is a string with no symbols. Empty string expressed by the symbol ε or ^ so that $|\varepsilon| = 0$. Empty string can be viewed as an empty symbol because both are composed of no symbols. Alphabet is the finite set of symbols

Grammar can be defined as a collection of sets of variables, terminal symbols, a symbol of the beginning, which is limited by rules. The example of automata can be seen at Figure 2.

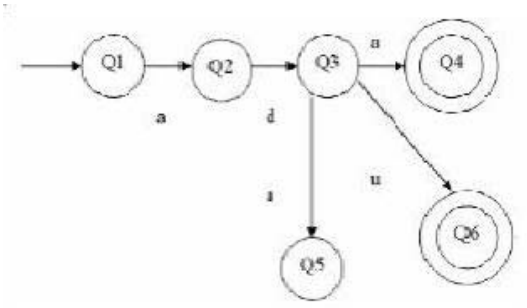


Figure 2. Sample of Automata

The rule of machine in Figure 2 is input string will be accepted if and only if it reaches the final state which is symbolized by a double circle. The machine has 6 states {Q1, Q2, Q3, Q4, Q5, Q6}. The machine has the initial state Q1. The machine has final state {Q4, Q6}. For example, if machine receives string "ada" or "adu", it will be accepted as correct input, but if machine receives string "add", it will be rejected.

4. JAVA CHARACTERS

4.1 Basic Java Characters

There are 20 basic Java characters, can be seen at Table 1 [4].

Table 1. Basic Java Characters

	Read		Read		Read		Read
	ha		da		pa		ma
	na		ta		dha		ga
	ca		sa		ja		ba
	ra		wa		ya		tha
	ka		la		nya		nga

4.2 Supplementary Characters

Supplementary characters in Java characters are used for several purposes. The first is to write vowels (i, u, é, e, o), because basic Java characters are actually syllables ended with 'a'. If the syllable begin with vowel, basic character 'ha' is used and combine with supplementary character. The second, supplementary is used to write some consonant in the end of word (r, h, ng) and to mark the end of the word if the word ended with consonant. Also, supplementary is used to write double consonant which one of the consonants is r or y (ex: pre, kya) and to write comma and period [4].

Table 2. Supplementary Characters

Symbol	Example	Read
		yi
		ye'
		ye
		yo
		yu
		yar
		yah
		yang
		k (consonant at the end of word)
		kra
		kre
		kya
		pa, (comma)
		pa. (period)

4.3 Consonant at the End of Syllable

Java character are actually syllables ended with 'a'. To write consonant at the end of syllable, there are symbols called '*pasangan*'. Each basic character has its own '*pasangan*' [4]. '*Pasangan*' has function to connect a closed syllable (syllable ending with consonant) to the next syllable.

For example, in the word 'banda' consists of two syllables: 'ban' and 'da'. To write the syllable 'ban', the character 'ba' should be written and followed by character 'na'. It would construct 'bana'. There is 'a' at the end of 'bana' This 'a' should be eliminate by using '*pasangan*' of the next syllable. So we should write '*pasangan*' 'da' instead of character 'da', to construct 'banda'. The '*pasangan*' 'da' has function to connect the closed syllable with the next syllable. It means the vowels followed the character will be turned off so that it becomes consonant.

All the basic characters have their respective '*pasangan*'. Some of these '*pasangan*' should be written below basic character and some of them should be written aligned with basic character. To understand more clearly how to write '*pasangan*', can look at Table 3 [4].

Table 3. *Pasangan*

Basic Char.	*Pasangan*	Read	Basic Char.	*Pasangan*	Read
		ha			pa
		na			dha
		ca			ja
		ra			ya
		ka			nya
		da			ma
		ta			ga
		sa			ba
		wa			tha
		la			nga

4.4 Numbers

Number symbols in Java character, can be seen at Table 4 [3].

Table 4. Number Symbols

Symbol	Num.	Symbol	Num.	Symbol	Num.
	0		4		8
	1		5		9
	2		6		10
	3		7		11

5. SYSTEM DESIGN

We design the Java characters word processing to be used with common Latin character keyboard. System will receive input from keyboard, and will generate the Java characters depend on the combination of Latin characters input. For example, if we press character 'm', followed by 'u' in keyboard, system will generate Java character for 'mu' (consist of basic character 'ma' and supplementary character 'u') . If we type sentence 'banda', system will generate Java character 'ba', 'na', and 'pasangan' 'da' .

Because not all Java characters could represent Latin character, the conversion should be used: Latin character 'z' will convert to 'j' in Java character, Latin 'q' will convert to 'k' in Java, Latin 'v' and 'f' will convert to 'p' in Java, and Latin 'x' will convert to 'ks' in Java (basic character 'ka' and '*pasangan*' 's'). To differentiate between e and **é** in Java character, the Shift key is used for **é**.

In system design, DFA (deterministic finite automata) is used to determine validity of input combination. The DFA design can be seen in Figure 3.

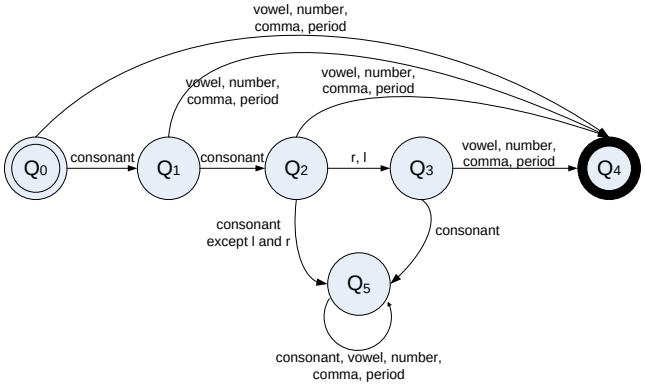


Figure 3. DFA Design

Input are consist of consonant (all alphabets except a,i,u,e,o), vowel (a,i,u,e,o), number (0..9), comma and period. Other input is discarded. Q_0 is beginning state of the system, if user input is consonant, it will go to state Q_1, but if user input is number, vowel, comma or period, it will go to accepted state Q_4. State Q_1 is similar to state Q_0, if user input is consonant, it will go to next state, but if user input is number, vowel, comma or period, it will go to accepted state Q_4. In state Q_2, the valid input only r, l, number, vowel, comma or period. If user input is r or l, it will go to state Q_3, and if user input is number, vowel, comma or period, it will go to accepted state Q_4. In state Q_3, valid input is only number, vowel, comma or period that will go to accepted state Q_4. Q_5 is error state. After system reach accepted state Q_4, system will execute rule to generate Java characters, but if system reach state Q_5, system will generate error message.

The rule to generate Java character is shown in pseudocode *GenerateChar* below. Input string *x* must be already passed the validation checking by DFA.

```
Procedure GenerateChar(string x);
//state Q0
if IsVowel(x[0])
    ShowCharacter('h'+x);
elseif IsNumber(x[0])
    ShowNumber(x);
elseif IsCommaPeriod(x[0])
    ShowCommaPeriod(x[0]);
else    // enter to Q1
{   if IsVowel(x[1])
        ShowCharacter(x);
```

```
Elseif IsNumber(x[1])
{   ShowCharacter(x[0]+'a');
    ShowNumber(x[1]);
} Elseif IsCommaPeriod(x[1])
{   if NoCharacterBefore()
    {   ShowCharacter(x[0]+'a');
        ShowCommaPeriod(x[1]);
    } Else
    {   ShowSupplementary(x[0]);
        ShowCommaPeriod(x[1]);
    }
} Else  // enter to Q2
{
    If IsVowel(x[2])
    {   if x[0]+x[1] in ['dh','ny','th','ng']
            ShowCharacter(x);
        Else
        {   ShowCharacter(x[0]+x[2]);
            ShowPasangan(x[1]);
        }
    } Elseif IsNumber(x[2])
    {   if x[0]+x[1] in ['dh','ny','th','ng']
        {   ShowCharacter(x[0]+'a');
            ShowNumber(x[2]);
        } Else
        {   ShowCharacter(x[0]+'a');
            ShowPasangan(x[1]);
            ShowNumber(x[2]);
        }
    }
    Elseif IsCommaPeriod(x[2])
    {   if NoCharacterBefore()
        {   if x[0]+x[1] in
            ['dh','ny','th','ng']
            { ShowCharacter(x[0]+x[1]+'a');
              ShowCommaPeriod(x[2]);
            } Else
            {   ShowCharacter(x[0]+'a');
                ShowPasangan(x[1]);
                ShowCommaPeriod(x[2]);
            }
        } Else
        {   if x[0]+x[1] in
            ['dh','ny','th','ng']
            { ShowSupplementary(x[0]+x[1]);
              ShowCommaPeriod(x[2]);
            }
        }
    } Else // enter to Q3
    {   if IsVowel(x[3])
        {   if x[2]='r'
            {   if x[3]!='e'
                { ShowCharacter(x[0]+x[3]);
                  ShowPasangan(x[1]);
                  ShowSupp_R();
                } Else
                { ShowCharacter(x[0]+'a');
                  ShowPasangan(x[1]);
                  ShowSupp_RE();
                }
            }
            Elseif x[2]='l'
            {   ShowCharacter(x[0]+'e');
                ShowPasangan(x[1]);
                ShowCharacter(x[2]+x[3]);
            }
        } Elseif IsNumber(x[3])
        {   if x[2]='r'
            {   if x[3]!='e'
                { ShowCharacter(x[0]+'a']);
                  ShowPasangan(x[1]);
                  ShowSupp_R();
                  ShowNumber(x[3]);
                } Else
                { ShowCharacter(x[0]+'a');
                  ShowPasangan(x[1]);
                  ShowSupp_RE();
                  ShowNumber(x[3]);
                }
            } Elseif x[2]='l'
            {   ShowCharacter(x[0]+'e');
                ShowPasangan(x[1]);
                ShowCharacter(x[2]+'a');
                ShowNumber(x[3]);
            }
        } Elseif IsCommaPeriod(x[3])
        {   if NoCharacterBefore()=false
            {   if x[2]='r'
                { ShowCharacter(x[0]+'a');
                  ShowPasangan(x[1]);
                  ShowSupp_R();
                  ShowCommaPeriod(x[3]);
                } Elseif x[2]='l'
                { ShowCharacter(x[0]+'e');
                  ShowPasangan(x[1]);
                  ShowCharacter(x[2]+'a');
                  ShowCommaPeriod(x[3]);
                }
            }
        }
    }
}
```

6. IMPLEMENTATION AND RESULT

We implemented word processing to be used in Windows operating system. Windows operating system have fonts For Java characters font using font creator software. We adjusted the fonts to be used with common Latin keyboard. The fonts will be used in word processing and should be installed first in Windows operating system.

In the word processing, there are three areas: Latin text area, Java text area, and wrong text area. These areas can be seen at Figure 4.

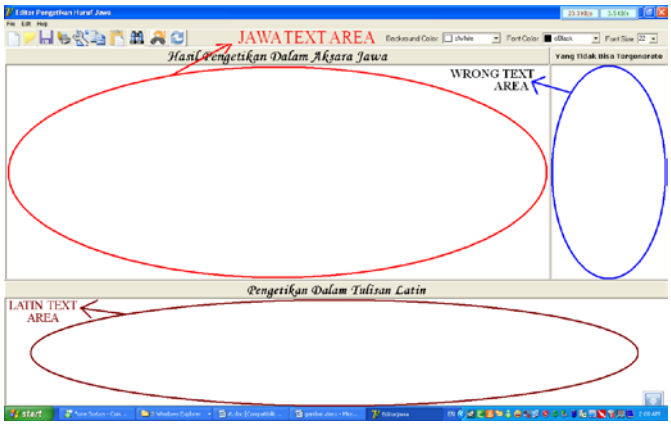


Figure 4. Areas in Word Processing

In Latin text area, user can type the Latin characters using keyboard, and the corresponding Java characters will be shown at Java text area. If the user typed a character that does not comply with the rules of Java word, then they will be displayed on the

wrong text area. For example, if user typed any characters that can't be translated to Java characters as seen at Figure 5, the characters will be shown at wrong text area.

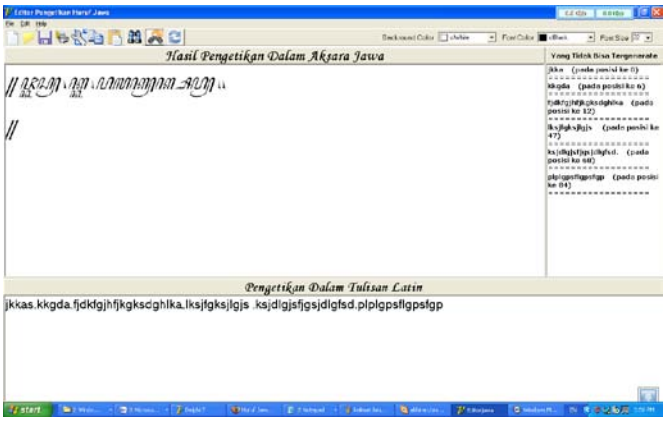


Figure 5. Writing any Characters

In this word processing, there are some standard features, such as copy/cut & paste, find & replace, change Java font color and font size and print.

The example of paragraph in the word processing can be seen at Figure 6.

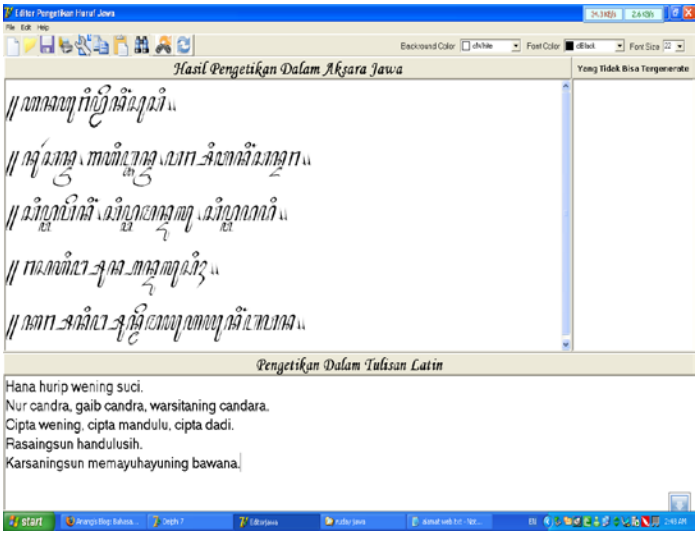


Figure 6. Example of Paragraph

7. CONCLUSION

We have developed Java character word processing that can be used to type Java character easily by using common Latin character keyboard. This word processing can display the combination of Java characters include basic, supplementary and '*pasangan*' characters, according to the typing keys on the keyboard, and has features like copy, cut, paste, find, replace and print. Using this word processing, people can convert Latin characters into Java characters even if they do not understand Java characters. This word processing could be further expanded to include another Java characters such as *murda*, *swara or rekaan* characters.

8. REFERENCES

[1] Eugnene X. 2004. Theory of automata, formal languages and computation. New Age International

[2] Barnes, Ian. 2006. Preservation of word processing documents. Australian National University

[3] Daryanto, 1999. Kawruh basa jawa pepak. Apollo, Surabaya.

[4] Tofani, Abi & Nugraha, Setyo. n.d. Tatanan anyar pinter basa jawi pepak. Pustaka Agung Harapan, Surabaya

The Design and Implementation of Digital Image Segmentation in HSV Color Space

Kartika Gunadi
Universitas Kristen Petra
Siwalankerto 121
Surabaya
(62) 031 2983000
kgunadi@petra.ac.id

Rudy Adipranata
Universitas Kristen Petra
Siwalankerto 121
Surabaya
(62) 031 2983000
rudya@petra.ac.id

Anthony Widianto
Universitas Kristen Petra
Siwalankerto 121
Surabaya
(62) 031 2983000

ABSTRACT
Segmentation subdivides an image into its constituent parts or objects[4]. The purpose of segmentation is to simplify or change from a digital image to the more meaningful and easier to analyze. Most of the segmentation process is done in RGB representation. In this paper, the segmentation process is done in HSV representation. The process consist several steps: changing the representation from RGB to HSV, search the centroid points using maximin algorithm, clustering by K-Means, and post processing of the results of clustering to eliminate noise and unnecessary detail. This application was created with C #. net with Microsoft Visual Studio 2005 as its IDE. The result indicates that the segmentation results are influenced by the centroid points. With determining the exact number of centroids, it gives clear segmentation result. Factors that can affect the centroid are intensity light and color range on the digital image.

Keywords
HSV, K-Means, Clustering, Image Segmentation.

1. INTRODUCTION
Segmentation is the process of digital image distribution into several parts. The segmentation function method of this feature is how to simplify or change the form of a digital image into meaningful and easier condition, so it can be analyzed faster depend on its application. The problems are how to recognize the objects or foreground image with the background setting and the fix parametric object segmentation as its real form.

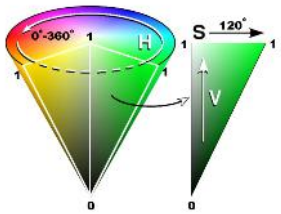


Figure 1. HSV Cone

Digital imaging is the group of color pixel that set on specific position and color to present a form or object. The pixel itself is well known as group of digital imaging represented by the value of RGB (*Red, Green,* dan *Blue*). The disadvantage of RGB representation itself can't to recognize the detail of highlights, shadows, and shading from a digital image. To complete the RGB method, the HVS digital image segmentation is added for excellent and perfect integration. HSV is a cylindrical coordinate representation function that used to rearrange RGB geometrical structure with relevant value, shown in Figure.1. This method is divided into few major topics that will be discussed bellow. Specific words may be known as the information follows.

Hue, is the basic color element to be used. Saturation is the perspective of lighting and the human sensitivity of color gradation variety depend on hue. Value is the intensity of color level or color brightness. HSV representation method provides the detail of color information for the human perspective about color and how the human use the color like an artist create a painting works.

The HSV per-pixel method use the technique of degenerate color that divided into two major function called color and gray histogram. The number of this method is set in a variety of description such as hue will set on [0^o,360 ^{o}], Saturation [0,1], and Value [0,1], so the quantization process for each dimension normalization can be counted as example bellow:

Quantization for Hue = 12^o, Saturation = 0.125, and Value = 0.125 it will be 30x8x8=1920 bin. If the amount of Hue Saturation is close to the gray color, the amount should be separated by the color and gray histogram for the perfect presentation. The detail of this function can be read as the table below.

Table 1. Correspondence of Histogram Bins and HSV Indices

Histogram bin	Range Parameter	Correspondence of Index
	v' = 0	
G(v)	atau	v = v'
	v' ∈ [1, 7] dan s' = 0	
	h' ∈ [0, 29]	h = h'
B(h,s,v)	*s'* ∈ [1, 7]	s = s'
	v' ∈ [1, 7]	v = v'

G(v) is represented the amount of pixel on gray histogram bin with the parametric described by (v) and B (h,s,v) represented the amount of pixel in color histogram bin with the parametric (h,s,v). So the total bin can be count 30 x 7 x 7 = 1470 bins.

Thus, the next step will continued with the process of clustering by the histogram value, labeling them on each pixel. The

clustering processes use the K-Means method to analyst all part of n clusters parts to “k” clusters observation based on mean or average value on close level. The value of “k” is counted by random method, commonly replacing with maximum heuristic algorithm histogram.

The clustering result is also describes as digital image segmentation, but it is need final process to eliminate the image noise and unnecessary pixel label. This process called post processing. The final result of all this process provides a new digital image that can be used for other application or processing and analyze as needed. The other function of this method (HSV segmentation) can be used also for object searching based on blob segmentation area. The widest and dominant blob will be describe as figure of background and the left blob (which has a different value of segmentation with the background) will be describe as foreground figure.

2. DESIGN AND IMPLEMENTATION

This Digital Image Segmentation in HSV Color Space contains 3 sections.

- Color space transformation.
  The function for transform RGB color space to HSV color space is defined by a set of equation number 1.

$$H1 = \cos^{-1}\left\{\frac{0.5[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\}$$

$$H = H1 \quad if \quad B \le G$$

$$H = 360° - H1 \quad if \quad B > G$$

$$S = \frac{Max(R,G,B) - Min(R,G,B)}{Max(R,G,B)}$$

$$V = \frac{Max(R,G,B)}{255} \qquad (1)$$

- Histogram HSV generation.
  Since the ranges of three dimensions in HSV color space are not the same (Hue: [0, 360◦], saturation: [0, 1], and value: [0, 1]), a quantization process is needed, shown in Figure.2. Quantization for Hue = 12°, Saturation = 0.125, and Value = 0.125. For correspondence of histogram bins, follow the Table 1.

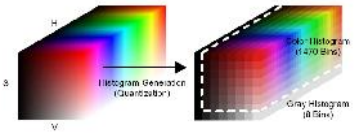


Figure 2. Histogram HSV *Generation* dan *Quantization*

- K-Means clustering and segmentation.
  K-Means clustering is a method for cluster analysis to observe many parts of *n* into *k* clusters or centroids where observations are based on the mean or average of the nearest neighbors. To find number of centroid from generated histogram, Maximin algorithm is used. From color and gray histogram bins, find the bin which has the maximum number of pixels to be the first centroid. For the remaining bins, calculate the minimum distance between it and its closest centroid. Bin which has a maximum value of the minimum distance is selected as the next centroid. Function for calculate the distance, is defined by

$$D_h^2(h_i, h_j^{(t)}) = \begin{cases} \left(\frac{360^0}{h_Q} - \left|h_i, h_j^{(t)}\right|\right)^2, if \left|h_i, h_j^{(t)}\right| > \frac{180^0}{h_Q} \\ \left(\left|h_i, h_j^{(t)}\right|\right)^2, otherwise \end{cases} \qquad (2)$$

Repeat the process until the number of centroid equals to KMax or the maximum value of the distance is smaller than a predefined threshold ThM. The KMax by default set by 10, based on the assumption that there should be no more than 10 dominant color in one image for highlevel image segmentation, and ThM = 25 is set empirically according to the human perceptions of different color in HSV color space.

After find the centroid, the next process is clustering the rest bins which not the centroid itself. The color and gray histogram will be clustered together. For color histogram, since Hue dimension is angle, the distance measurement between a color histogram bin vector $\mathbf{B}_i = (h_i,s_i,v_i)$ with a cluster centroid vector $C_j^{(t)} = (h_j^{(t)},s_j^{(t)},v_j^{(t)})$ in the current iteration t is defined in the form of the Euclidean distance.

$$D^2(B_i, C_j^{(t)}) = D_h^2(h_i, h_j^{(t)})^2 + (s_i - s_j^{(t)})^2 + (v_i - v_j^{(t)})^2 \qquad (3)$$

Each color histogram bin clustered to nearest centroid by the distance measurement.

$$\phi_{(j|B_i)}^{(t)} = \begin{cases} 1, if \ \ j = \arg\min_k D^2(B_i, C_k^{(t)}) \\ 0, otherwise \end{cases} \qquad (4)$$

For the gray histogram bins since no Hue value information, the distance measurement between a gray histogram bin $\mathbf{G}_i = (v_i)$ with a cluster centroid vector Cj(t) = (hj(t),sj(t),vj(t))

$$D^2(G_i, C_j^{(t)}) = (s_j^{(t)})^2 + (v_i - v_j^{(t)})^2 \qquad (5)$$

After all color and gray histogram bins clustered to nearest centroid then update centroid. Again, since the hue dimension is circular, the indices in the hue dimension should be considered not absolutely but relatively. The method to calculate the relative value of the original Hue index h_i to the centroid $C_j^{(t)} = (h_j^{(t)},s_j^{(t)},v_j^{(t)})$

$$\bar{h}_{i,j}^{(t)} = \begin{cases} h_i - \frac{360^0}{h_Q} \ if \left|h_i - h_j^{(t)}\right| > \frac{180^0}{h_Q} and \ h_j^{(t)} < \frac{180^0}{h_Q} \\ h_i + \frac{360^0}{h_Q}, if \left|h_i - h_j^{(t)}\right| > \frac{180^0}{h_Q} and \ h_j^{(t)} > \frac{180^0}{h_Q} \\ h_i \quad , \quad otherwise \end{cases} \qquad (6)$$

Then update all dimension of centroid according to the following equations. Where $B(B_i)$ is the number of pixels in the color histogram bins for histogram bin vector B_i and $G(G_i)$ is the number of pixels in the gray histogram bins for histogram bin vector G_i.

$$h_j^{(t+1)} = \frac{\sum_{i=1}^{N_B} \tilde{h}_{i,j}^{(t)} \, \phi_{(j|B_i)}^{(t)} \, B(B_i)}{\sum_{i=1}^{N_B} \phi_{(j|B_i)}^{(t)} \, B(B_i)},$$

$$s_j^{(t+1)} = \frac{\sum_{i=1}^{N_B} s_i \, \phi_{(j|B_i)}^{(t)} \, B(B_i)}{\sum_{i=1}^{N_B} \phi_{(j|B_i)}^{(t)} \, B(B_i) + \sum_{i=1}^{N_G} \phi_{(j|G_i)}^{(t)} \, G(G_i)},$$

$$v_j^{(t+1)} = \frac{\sum_{i=1}^{N_B} v_i \, \phi_{(j|B_i)}^{(t)} \, B(B_i) + \sum_{i=1}^{N_G} v_i \, \phi_{(j|G_i)}^{(t)} \, G(G_i)}{\sum_{i=1}^{N_B} \phi_{(j|B_i)}^{(t)} \, B(B_i) + \sum_{i=1}^{N_G} \phi_{(j|G_i)}^{(t)} \, G(G_i)},$$

After all centroid has been updated, check whether the clustering process is converged according the summation of distances between each histogram bin to nearest cluster centroid.

$$\Delta^{(t)} = \sum_{i=1}^{N_B} \sum_{j=1}^{K} \phi_{(j|B_i)}^{(t)} \, D^2(B_i, C_j^{(t)}) \, B(B_i) + \sum_{i=1}^{N_G} \sum_{j=1}^{K} \phi_{(j|G_i)}^{(t)} \, D^2(G_i, C_j^{(t)}) \, G(G_i) \quad (8)$$

If the difference of total distortion $|\Delta(t+1) - \Delta(t)|$ is less than predefined threshold. If not converged then repeat the clustering process until converged or when maximum iteration is reached. After the clustering process is completed then the digital image has been labeled or segmented in the HSV representation.

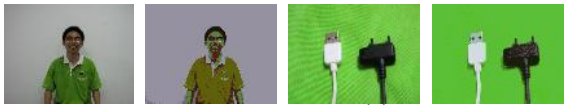
In addition, a filter for post-processing is introduced to effectively eliminate noise (small labeled). And also suggest the candidate of object from the segmentation result, background or foreground.

- Post processing.
  To remove the noise, common Spatial Filters like Median and Mean Filter are applied to labeled image.
- Suggest the candidate of object.
  The first step on seeking object candidate is starting with the label image of the background candidate. The initial process starting with counting the roving image, the label which has the maximum amount will be set as background. Wide of the initially process depend on the user input where the minimal wide per pixel will be affected through initial system. When the process is finished by the background labeling, the seeking object process will be continued as further step.
  In assumption, the left labeled image exclude the background will be initialed as foreground image. This process starts by the verification loop that allows the system to count the process with valid value. The group of the left labeled image will be set partly as blob. This blob will be categorized into few threshold images, then selecting by user input. When the large of the blob is low from the threshold, it will be vice versa and will be used as object candidate.

3. EXPERIMENTS

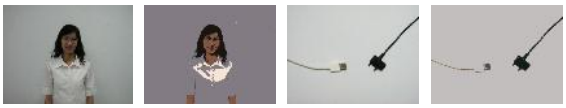
The experiments are performed in Intel® Core i5™ 430M 2 core @ 2.8 GHz, RAM 4 GB, and graphic card ATI Mobility Radeon HD 5470 1 GB VRAM. Digital images are captured by pocket digital camera Canon PowerShot™ A430 4.0 Megapixels.

- Digital image background experimental factors.
  The process of testing is done by giving the input image with a background color must be contrast and had a different color contrast with the object itself. In this test, a resolution of 1024x768 pixels is selected and the parametric segmentation mode is set by default.



(a) (b) (c) (d)

Figure 3. Background contrast with the object: (a) Original 1, (b) Segmentation result of (a) (k=10, time=6029 ms), (c) Original 2, (d) Segmentation result of (c) (k=9, time=2465 ms)



(a) (b) (c) (d)

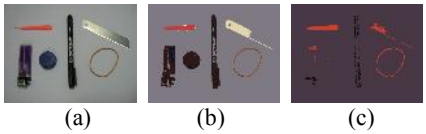
Figure 4. Background not contrast with the object: (a) Original 1, (b) Segmentation result of (a) (k=10, time=5647 ms), (c) Original 2, (d) Segmentation result of (c) (k=14, time=9172 ms)

Image which has a contrast background, give more clearly segmentation result. Centroid searching parametric testing.

The testing process of the centroid searching parametric is done by giving the value of the parameter range K-Max which limit the number of centroid and the threshold parameters that limit the calculation. The experiments also distinguished between the maximin algorithm automatically searches and manually from users inputs where the choice is limited by human perception about colors and objects. In this test, a resolution of 1024x768 pixels is selected and the parametric segmentation mode is set by default except for the centroid search, shown in Figure.5 and Figure.6.

(a) (b) (c) (d)

Figure 5. Maximin algorithm: (a) Source Digital Image, Segmentation result: (b) K-Max default (100, (c) K-Max set 6 (d) K-Max default (100)



(a) (b) (c)

Figure 6. Manually user inputs: (a) Source Digital Image, Segmentation result: (b) Try I: K-Max set 7 (candle, spidol, rubber, bottle cap, lighter, razor blade, and

background, (c) Try II: K-Max set 7 (candle, spidol, rubber, bottle cap, lighter, razor blade, and background)

It can be seen that finding centroid using maximin algorithm with various range K-Max and Threshold give different number of centroid found. Meanwhile, user manual input may affected a different segmentation results, though the selecting object testing process had been used twice. This may be happened because the value of the chosen centriod HSV may be different in the same object, for example the gradation looks like the same but actually has different value.

- Flux testing factors.

The Flux testing factor process is accomplished by providing a digital image that has a range of composition and same object but different in flux. Flux settings not from digital image reprocessing, but from exposure settings on the camera, is raised, lowered, or normal. In this test, a resolution of 1024x768 pixels is selected and the parametric segmentation mode is set by default.

Figure 7. (a) Exposure normal. (b) Segmentation result (12 centroid finded)

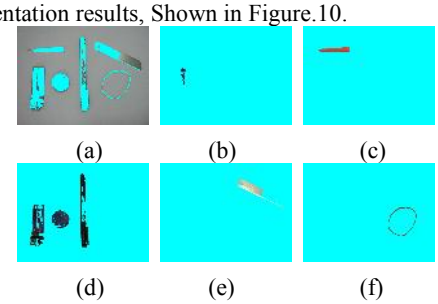
Figure 8. (a) Exposure -2. (b) Segmentation result (11 centroid finded)

Figure 9. (a) Exposure +2. (b) Segmentation result (13 centroid finded)

It can be seen that lower or more exposure make digital images loose of color information. Lower exposure makes color fades to dark and more exposure makes color fades to white. Various flux not affected the segmentation process, but affected the segmentation result because the finded centroid are different.

- Testing the segmentation result to find candidate object

The testing process to find candidate object from segmentation result are performed by providing different kinds of segmentation results, Shown in Figure.10.



(a) (b) (c)

(d) (e) (f)

Figure 10. Source image Fig. 10 (a) with 6 centroid, (a) candidate background, (b) candidate foreground 1, (c) candidate foreground 2, (d) candidate foreground 3, (e) candidate foreground 4, (f) candidate foreground 5

(a) (b) (c) (d)

Figure 11. Source image Fig. 5 (a) with 4 centroid, (a) candidate background, (b) candidate foreground 1, (c) candidate foreground 2, (d) candidate foreground 3

It can be seen that segmentation result affected the process to finding the candidate object. When the results of the segmentation equal to the expectation, the search of object candidate will be valid.

4. CONCLUSION

Based on the test results, it can be inferred conclusions as follows:

- Segmentation results is only affected by the centroid values and number of centroid that is used in the K-Means clustering process. The amount processing depends on the centroid size.
- Within searching process, the number of objects in the centroid digital image does not affect to the found of centroid. The process itself will only be affected by the HSV values information on various objects in the digital image.
- Light intensity/flux is affected to the clarity information on digital image where this information would be a reference in the selection process of centroid. The image that gets a high light intensity, will be resulting to digital image tends towards white color when compared with normal intensity. So do the opposite conditions, when the image gets a low intensity, the digital images tend to black. This is similar to the subject of human eyes to see objects. When the light intensity is too high or low, then the information can be captured by eyes is reduced or absent.
- The contrast background with the object on the digital images will produce a clear segmentation between the background and the object. Meanwhile, not contrast background makes some or all of the segmentation on the object into one cluster with a digital image background.
- The right determination of the exact amount of centroid, resulting clear and fixed segmentation. The large number of the centroid makes the segmentation became scattered and produce some noise.
- The lowest number of centroid, will only make the segmentation results looks like forced.

5. REFERENCES

[1] Adipranata, R., Handojo, A., Prayogo, I., Yuliana, O.Y. (2005). Perancangan dan Pembuatan Aplikasi Segmentasi Gambar Dengan Menggunakan Metode Morphological Watershed. *Jurusan Teknik Informatika, Universitas Kristen Petra*. Retrieved Oktober 28 2010, from http://fportfolio.petra.ac.id/user_files/94-014/Perancangan%20dan%20Pembuatan%20Aplikasi%20Segmentasi%20Gambar%20Dengan%20Menggunakan%20Metode%20Morphological%20Watershed.pdf

[2] Chen T.W., Chen, Y.L., and Chien, S.Y. (2008*). Fast Image Segmentation Based on K-Means Clustering with Histograms in HSV Color Space. Media IC and System Lab*

Graduate Institute of Electronics Engineering and Department of Electrical Engineering National Taiwan University, Taipe. Retrieved September 25 2010, from http://media.ee.ntu.edu.tw/~variant/paper/mmsp08_seg.pdf.

[3] *Color Models*. (n.p). Retrieved October 28 2010, from http://www.photozone.de/colorimetric-systems-and-color-models.

[4] Gonzalez, Rafael C., and Woods, Richard E. (2002). *Digital Image Processing Second Edition*. New Jersey: Prentice Hall.

[5] Padilla, M., and Fan, Z. (2003). *Automatic Face Detection Using Color Based Segementation and Template/Energy Thresholding*. *Department of Electrical Engineering, Stanford*. Retrieved September 25 2010, from http://www.stanford.edu/class/ee368/Project_03/Project/reports/ee368group16.pdf.

[6] *RGB Cube to HSV Cone*. (n.d). Retrieved October 28 2010, from http://www.flashandmath.com/advanced/color/.

[7] Sural, S., Qian G., and Pramanik S. *Segmentation and Histogram Generation Using the HSV Color Space For Image Retrieval*. *Dept. of Computer Science and Engineering, USA*. Retrieved September 25 2010 , from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.9445&rep=rep1&type=pdf.

[8] Wang, Yanjiang and Yuan, Baozong. (2001). *A novel approach for human face detection from color images under complex background*. *IEEE Computer Society*, 34, 1983-1992.

Interlace and De-interlace Application on Video

Liliana
Informatics Department, Faculty of Industrial Technology, Petra Christian University
Surabaya, Indonesia
lilian@petra.ac.id

Justinus Andjarwirawan
Informatics Department, Faculty of Industrial Technology, Petra Christian University
Surabaya, Indonesia
justin@petra.ac.id

Gilberto Erwanto
Informatics Department, Faculty of Industrial Technology, Petra Christian University
Surabaya, Indonesia

ABSTRACT

Media is developed as a medium of exchanging data demands speed and data accuracy. Both of these need for greater storage and bandwidth as a challenge. While a large storage and bandwidth costs are greater. To face the challenges that exist, compresion and interlace video techniques are most often used. Both are at risk video quality decrease but can achieve the desired results.

In this research, we use interlace as a solve method in our developed application. The goal of our application is to deliver a video file through a limited bandwith. Instead passing the fully video, it will deliver a part of the file. Before display the video, de-interlace method will manipulate the delivered file and reconstruct the not-delivered part.

Test results indicate that the client does not affect the amount of bandwidth consumption during multicast, but the file format. Smoothness of multicast depends on several factors like the type of image compression, including interlaced video. In several cases, when the delivery delay time is big enough, deinterlace method cannot reconstruct the image fully.

Keywords

Interlaced Video, Deinterlacing, bandwidth, compression

1. INTRODUCTION

In recent years, the speed and accuracy in a data access through a medium is very important, especially via internet. Frequently accessed data video is displayed through streaming or buffering. In the process of streaming video through the Internet, the

buffering process will download data before being displayed on the user. Buffering data before downloaded by a user is placed on the memory area. The major challenge to be faced is the video data size is large enough for loading images as well as sound contained in it.

If a user requests a video on a web page, such as Youtube, Youtube will take the data source from the server and will be displayed on the client's screen. However, because of the limitations of hardware devices that are owned by the user, such as bandwidth limitations, will make the data load process longer. Data load process with a very large size, will require greater bandwidth.

There are several ways to speed up the process of sending it. Firstly, compression, it will reduce the physical size (the size or color quality on the video) before sending data. Secondly, the video delivery is resolved into two parts and sent one of them. This is called interlace method. For the data is transmitted partially, the resulting quality is also only a part of it. There are several techniques to manipulate the transmitted data. De-interlace is one of those[1],[2],[5].

Interlace method is often seen in the information conveyed through the television where noise is often seen on the displayed image. Therefore, the purpose of this study is to create an application that can process video data in order to produce a video that has a smaller size than the original data, much faster in delivering data, but still maintain the existing quality of the original video data.

2. AVI (Audio Video Interleave)

AVI is one of interlace methods. It breaks audio and video streams into several parts. Each part is arranged so it can represent its neighbor parts which appear at each time period. This regulation allows linear reading along the video playback, but doesn't allow writing whole file before displaying it one by one. There are two video displaying methods, progressive and interlace scan.

2.1 Progressive scan

It is a method for displaying each frame of a video sequently. Each appearance of the frame shown in full image without any removal field. Scanning using this method is usually done at 50 fps. At 50 fps motion can produce the perfect motion sequence, the high temporal and vertical resolution[1],[2].

2.2 Interlace scan

It is a method for displaying a video which only shows a part of an image on every frame of it. This method is the most commonly used in television[1]-[2]. Interlacing is done by only displaying either even or odd rows, or diagonally. Interlace scan often shows the effect of horizontal lines, especially on the moving object. The effect is called mice teeth combing effect, see Fig 1. This effect is sometimes seen clearly, but sometimes it is almost invisible.

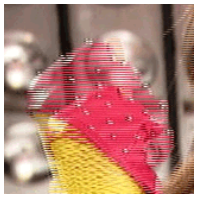


Figure 1. Mice Teeth effect from Interlaced Video

source:http://casualmuxer.wordpress.com/2008/07/22/sekilas-tentang-interlaced-video/

In the interlaced video, combing effect caused by the woven of the two fields becomes one whole frame, see Fig 2. A field has half the vertical resolution (spatial) of a frame. An interlaced video with a resolution of 640 × 480 pixels at 30 frames per second (fps) has sixty 640 × 240 fields per second. To render the fields had become one whole frame, each of the two fields are woven by put the pixel rows are alternately into one frame. Weaving process is the cause the effect of combing[5].

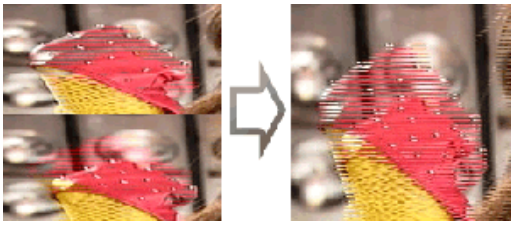


Figure 2. Combining 2 Field become 1 Frame

source:http://casualmuxer.wordpress.com/2008/07/22/sekilas-tentang-interlaced-video/

3. DE-INTERLACE

De-interlace is a converting process of interlaced video which composed of a set of half part images to be whole image (non-interlaced). Several types of methods are commonly used are blend, bob, discard, linear, doubler, mean, motion blur, and motion compensation.

Blend method is often used on the TV screen. De-interlacing blending produces a little bit blur effect. Blending method performs mixing between the simulated and motion blur as well as combining two consecutive images together sequently. This method reduce the quality a quarter of the original quality [1]-[4].

Area-based methods do not blend in all the parts, but only in the mice teeth area. This can be done by comparing the frame by time or position. This way will give good results in scenes that do not involve a lot of movement, so, no blur is produced.

Another frequently used method is the method doubler. After the disposal of one field, e.g. the even lines, then on even-numbered rows are filled with previous odd lines copy value. This method is often used because it does not require a long processing time.

Interpolation method calculates the mean or average of the remaining neighbor fields[5]. For example, there is an image resolution of 5 x 5 as in Fig 3 (left). From the top to bottom are line [0..4] and from left to right are columns [0..4]. If interlace even fields only is applied, then the even lines will be eliminated, and the rest is just odd lines will produce an image as shown in Fig 3 (middle). If diagonal interlace is applied, then it will produce Fig 3 (right), where 'X' states the removing data, and O states remaining data.

	0	1	2	3	4
0	O	O	O	O	O
1	O	O	O	O	O
2	O	O	O	O	O
3	O	O	O	O	O
4	O	O	O	O	O

	0	1	2	3	4
0	X	X	X	X	X
1	O	O	O	O	O
2	X	X	X	X	X
3	O	O	O	O	O
4	X	X	X	X	X

	0	1	2	3	4
0	X	O	X	O	X
1	O	X	O	X	O
2	X	O	X	O	X
3	O	X	O	X	O
4	X	O	X	O	X

Figure 3. Original image (left), even field only (centre), diagonal interlace (right)

4. GENERAL SOFTWARE SYSTEM

This application is divided into two sections, client interface as a medium for displaying results of interlaced video, and server as data processing before data is sent. The outline of the system software can be seen in Fig 4.

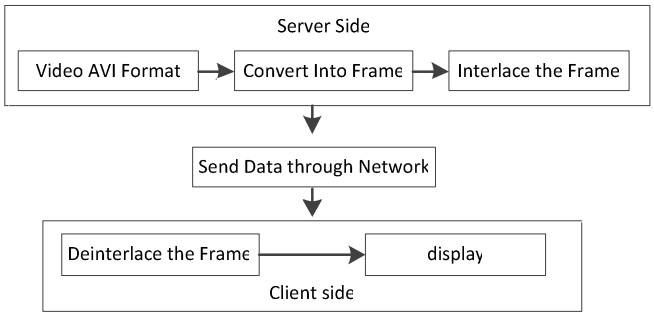


Figure 4. General Software System

4.1 Server System

In the server side, there are several important steps that must be done. There are three main processes on the server, the Open Video Publication, Video Registration, Publication and Video.

Open Video Publication

The first time the server runs, the Open Publication Video will be done. The server chooses a video file which will be published. Then the server initiates the publication of the desired video. The server will send information about the video which will be published as well as the server IP multicastly to all active clients. Data contain author, title of video, ports, and video address.

Registration Video

In the video the registration process, the server receives the IP of the client listed in the published video group. Video's information will be displayed on the user screen. Beside that, the server can perform other activities, multicast features. Through these features the server sends the output of send-image to the client that has been registered earlier. When the client receives the IP of the user who access the video, the server will perform its data storage. After storing the data, the server can perform multicast to do at any time, not only occur during multicast.

Video Publication

In the process of publication of this video, the server sends the result of the converted video image. Before the image is sent to

the client, the image will be processed first, using the interlace method. Then server multicast to all clients connect to the server.

Interlace Scanning Frame

At this stage it will be scanning the frame diagonally. Scanning process will be done starting from the top left corner to the right-moving frame. After arriving at the border width or horizontal image frame, it will form a line, then going back to start scanning from the left under the line that had been made earlier. Flowchart of Frame Interlace Scanning can be seen in Fig 5.

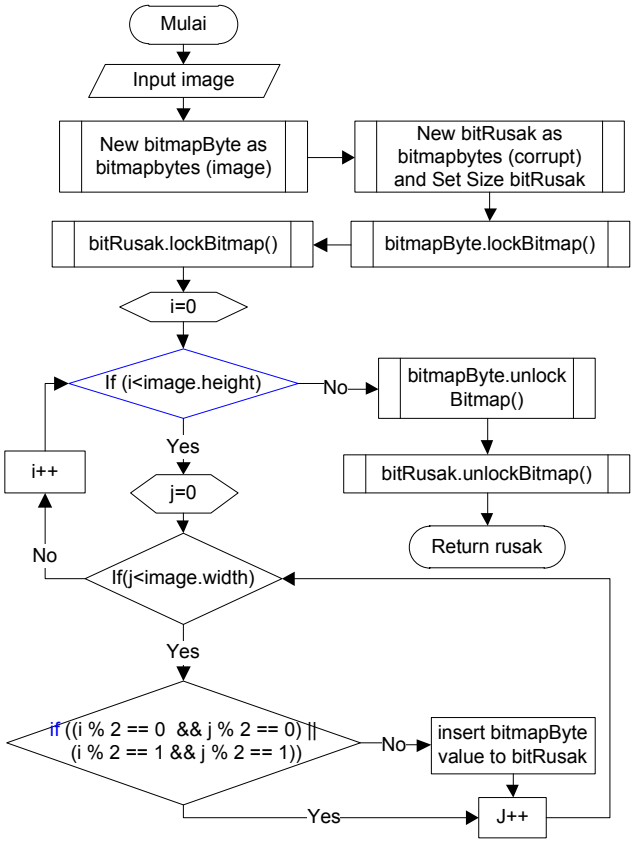


Figure 5. Interlace Scanning Frame Flowchart

4.2 Client System

Client side has many features, Refresh, Join Video, Pause, Download Video. General flowchart of the server side is showed in Fig 6.

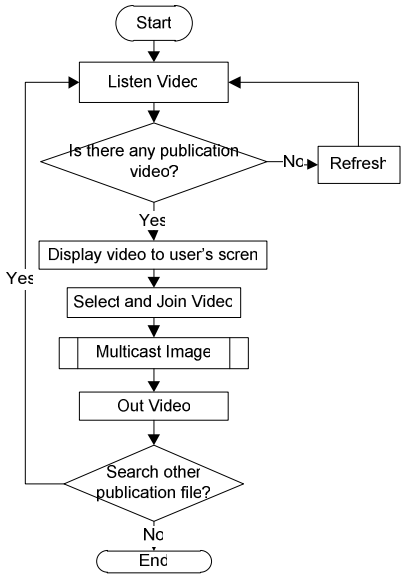


Figure 6. Interlace Scanning Frame Flowchart

Multicast Image

The multicast process starts by setting up image image stream through a socket for receiving the transmitted image server. Once the image is received, the image data stream is not directly displayed on the screen. It would be processed first, which performed in order to regenerate normal size image and the fulfil particular pixel on the image which is remain, processed by interpolation de-interlace or mean. Flowchart multicast image can be seen in Fig 7.

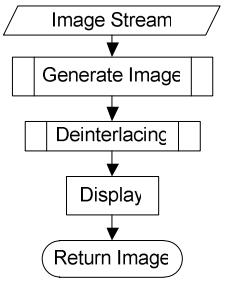


Figure 7. Multicast Image

Generate Image

This proses produces whole size image and then entry the pixel value from bitRusak to the new image. The complete process can be shown in Fig 8.

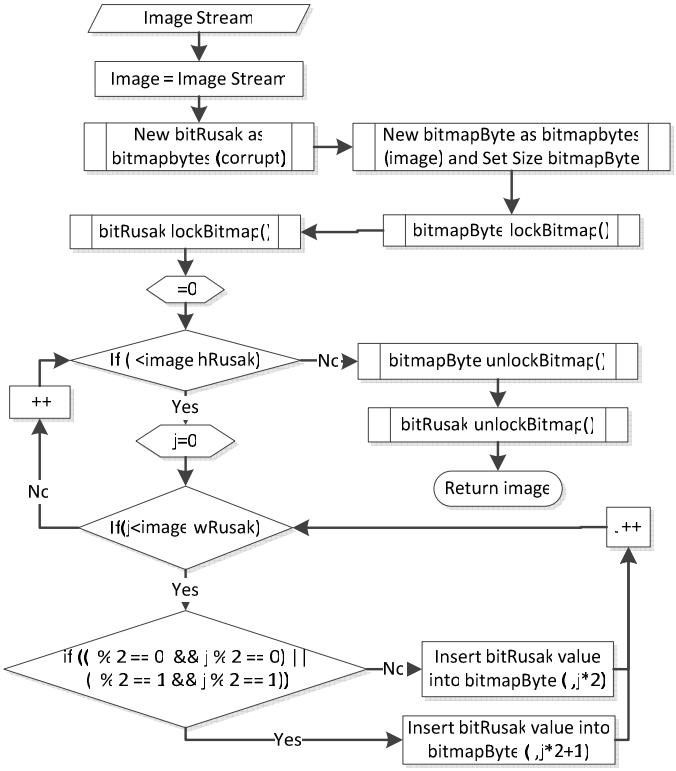


Figure 8. Generate Image Process

De-interlacing

Before displaying the image to screen, it will apply de-interlace method which is shown in Fig 9.

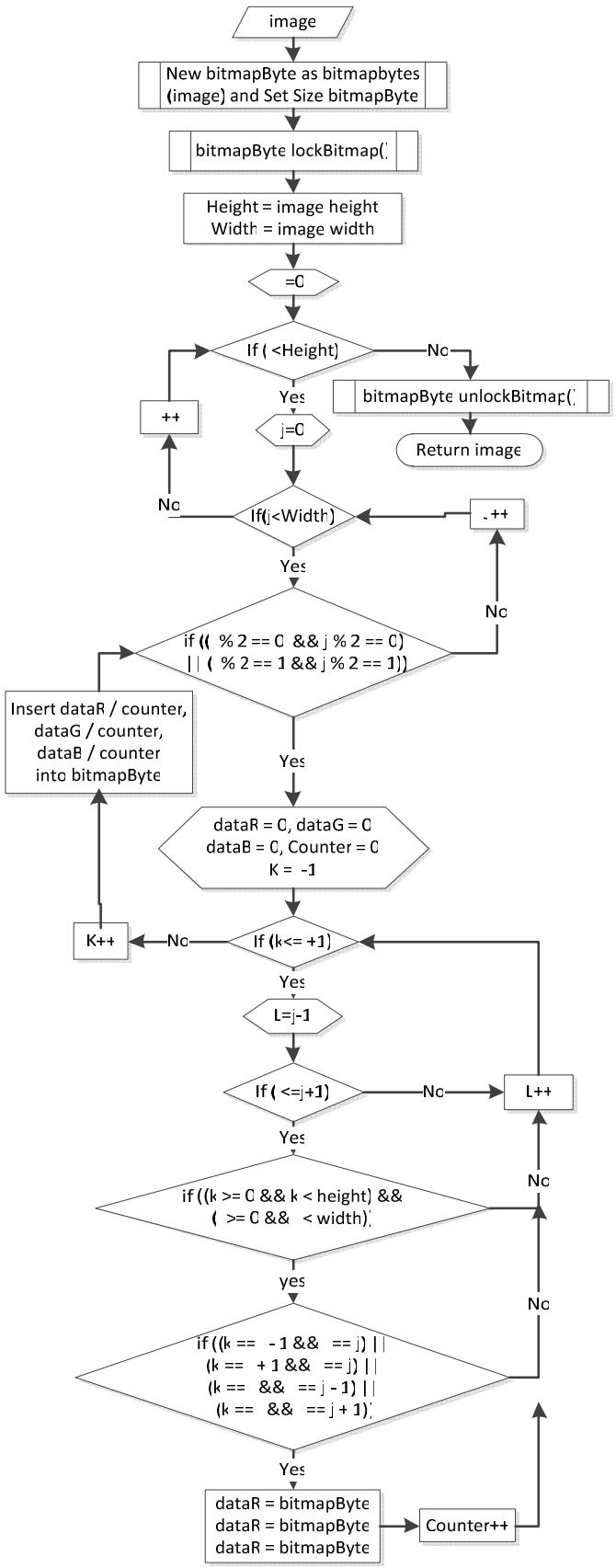


Figure 9. Deinterlacing Proses

5. EXPERIMENTS

Experiments are conducted to determine the delay generated by each image format. We use four kinds of image format, JPEG, BMP, GIF, and PNG. Testing is performed by comparing the application using the interlace method and without interlacing. Testing is done by taking the initial 10 frames are displayed, the calculation is based on the difference in time required to display each frame received. Testing will calculate the time difference between the current frame and the next appearance on a client. The experimental results using the interlacing and the delay that occurred is displayed in Table 1 and Table 2, while those without interlace shown in Table 3 and Table 4.

Table 1. Time at taking frame using interlace

Frame	JPEG	BMP	GIF	PNG
1	7949030	8301421	8891432	8798425
2	7949514	8305212	8892275	8799766
3	7949982	8309050	8893164	8800578
4	7950372	8326319	8894007	8801748
5	7950793	8333994	8894880	8802902
6	7951183	8455924	8895723	8803791
7	7951636	8482070	8896596	8804587
8	7952041	8549369	8897485	8805570
9	7952447	8562458	8898000	8806490
10	7952884	8588650	8898375	8807410

Table 2. Delay frame produced from tabel 1

Frame	JPEG	BMP	GIF	PNG
1-2	0.484	3.791	0.843	1.341
2-3	0.468	3.838	0.889	0.812
3-4	0.390	17.269	0.843	1.170
4-5	0.421	7.675	0.873	1.154
5-6	0.390	121.930	0.843	0.889
6-7	0.453	26.146	0.873	0.796
7-8	0.405	67.299	0.889	0.983
8-9	0.406	13.089	0.515	0.920
9-10	0.437	26.192	0.375	0.920
Average	0.428	31.914	0.771	0.998

Tabel 3. Time at taking frame without interlace

Frame	JPEG	BMP	GIF	PNG
1	9305662	9916094	9854083	9602984
2	9305724	9921211	9854520	9604310
3	9305834	9923769	9854988	9605699
4	9305943	9949431	9855441	9607103
5	9306052	10159923	9855877	9608460
6	9306130	10252510	9856299	9609380
7	9306208	10270544	9856704	9610332
8	9306286	10278250	9857110	9611330
9	9306380	10373224	9857531	9612329
10	9306442	10378372	9857952	9613202

Tabel 4. Delay frame produced from Tabel 3

Frame	JPEG	BMP	GIF	PNG
1-2	0.062	5.117	0.437	1.326
2-3	0.110	2.558	0.468	1.389
3-4	0.109	25.662	0.453	1.404
4-5	0.109	210.492	0.436	1.357
5-6	0.078	92.587	0.422	0.920
6-7	0.078	18.034	0.405	0.952
7-8	0.078	7.706	0.406	0.998
8-9	0.094	94.974	0.421	0.999
9-10	0.062	5.148	0.421	0.873
Average	0.087	51.364	0.430	1.135

From table 2 and table 4, the result of compressed videos such as JPEG and GIF are not suitable with this method. It is because of the compact information from the compressed file. It will take no time delay in delivering the file. While the uncompressed video, such as BMP and PNG, this method is well done. This method will make a significant result as expected.

6. CONCLUSION

Based on the test results can be summed up some of these following:

- Image format types affect the amount of data loss in the event of multicast. Type JPEG format has a lower data rate than the three types, e.g BMP, GIF, and PNG. Raw format file consumes a large time in delivery. It makes our system reconstruct the full image with only a part of the image.
- The type of image format type affects delay in the event of multicast. BMP format type provides the greatest delay compared to other types, JPEG, PNG and GIF. Selection of the proper type of format is recommended to reduce the delay in the event of multicast.

- Diagonal interlace method has less degradation than the odd events interlace method.
- The decline is caused by interlace method is not very significant, considering the parts are sent only half of the original image.

7. REFERENCES

[1] G. de Haan, E. Bellers. 1998. De-interlacing: An overview. Proceedings of the IEEE, 86, (9), 1839–1857.

[2] G. de Haan, R. Lodder. 2002. De-interlacing of video data using motion vectors and edge information. Digest of the ICCE'02, 70–71.

[3] H.S. Oh, Y. Kim, Y. Jung, A. W. Morales, and S.J. Ko. 2000. Spatio-temporal edge-based median filtering for deinterlacing. In ICCE. 2000 Digest of technical papers. International Conference on Consumer Electronics, pages 52 – 53.

[4] M. Zhao, G. de Haan. 2004. Intra-field deinterlacing with advanced up-scaling methods. Proc. of the ISCE, Reading, UK, 315–319.

[5] Interlaced Video. 2011. Sekilas tentang interlaced video. Retrieved Oct. 8, 2011, from http://casualmuxer.wordpress.com/2008/07/22/sekilas-tentang-interlaced-video/

Contacts Backup Management on Cellular Phones

JustinusAndjarwirawan
Informatics Department of
Petra Christian University
Siwalankerto 121-131, Surabaya
+62816519921
justin@petra.ac.id

Andreas Handojo
Informatics Department of
Petra Christian University
Siwalankerto 121-131, Surabaya
+62816538522
handojo@petra.ac.id

Angela Feliciana Soenjaya
Informatics Department of
Petra Christian University
Siwalankerto 121-131, Surabaya
+6281703483022
fei_en@yahoo.com

ABSTRACT
The development of time and technology has changed social lifestyle nowadays. Cellular phone has become the communication life style in most society. The problem appears when the cellular phone's contacts list is lost. It is quite difficult and needs more time to restore. Back-up and synchronizing contact may overcome this problem.

In this paper, the discussion covers back-up and synchronizing contacts which is performed by an application together with a server and Internet connection. A cellular phone user must connect to an account in a specific website and install application on the cellular phone. Then the users may use the features provided by application which builds on Java ME. First, the application retrieves all the contacts list from the database and enable the users to mark the data before the back-up process performed. After the application downloads the contacts data from the server, synchronizing process is performed.

The result of the experiments shows that this application works on Nokia feature phones with operating system S40 6th Edition and S60, also on Sony Ericsson with Java Platform 8 and latest. This application works well when supported by the availability and stability of Internet connection with compatible cellular phone specification.

Keywords
Personal Information Management, Contacts Back-up, Contacts Restore, Synchronize.

1. INTRODUCTION
Mobile phones have become something that is mandatory for most people to have. This is related to lifestyle trends and the need for mobility. Since it was first developed, features that are still needed in the mobile phone is to call and send short messages known as SMS (Short Messaging Service).

The second feature on the mobile phone is often used by users to continue to keep in touch with their contacts. Users need to type the destination number to call or send SMS. In fact, not just one or two people's phone numbers who should be remembered. This human limitationis an idea to mobile phone developers to add contact phonebook feature. With this feature on the mobile phone users can keep a list of names and phone numbers like a phone book. This facilitates the user's mobile phone to contact people who want to be contacted. Increasingly, phonebook storage capacity becomes greater.

One of the things that matter to most people when something happens is lost list of phone numbers stored in their phonebooks. It can also be quite a hassle when replacing with a new mobile phone, how to move the list of phone numbers from the older phone. Time consuming and sometimes difficult to check any of the numbers that have been removed and which ones have not. As a result, there are omissions or stored twice.

Needed a way to back-up data phone number which can be accessed at any time when needed. Mobile phone users can upload data on a contact to a server via a website.

Connection to the Internet must be done through the mobile phone itself. An application on the mobile phone reads the data contact and pick it for upload. File was uploaded successfully stored in the database with a specific format. Users can download the data again. Successful data can be downloaded directly used without having to bother inputting the application due to the cellular phone earlier. Such applications are expected to answer the most frequently encountered problems associated with the use of the features of the phone book on your mobile phone.

2. PIM IN JAVA ME
Mobile phones generally have a phone book feature to store phone numbers. This information includes personal information that cannot be accessed directly by the MIDP (Mobile Information Device Program) [1] application on Java runtime environment phones. This personal information can be accessed and managed using the PIM (Personal Information Management) [2]API (Application Programming Interface) for Java ME (Micro Edition) [3]. Other personal information isaccessible to the calendar which is used to store important events and also the to-do lists.

Figure 1 shows the hierarchy of the major classes and interfaces in the PIM API.

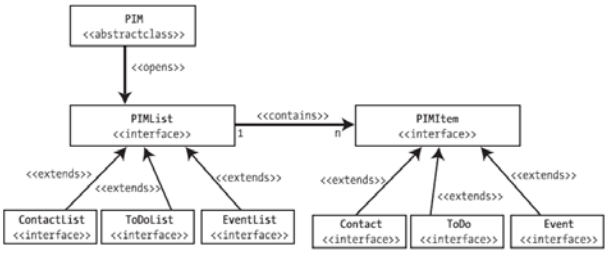


Figure 1. Major classes & Interfaces of the JSR 75 PIM API

JSR (Java Specification Requests) 75 is a Java API package to handle PIM accesses on Java ME enabled mobile phones.

PIM is a singleton object used to enter the lists and records. PIM represents a database of listsof items or records. Lists and the itemsare accessed via the interface and PIMItemPIMList interface.
Before using it, should be checked first whether the PIM Optional Package is available on the mobile phone:

```
String currentVersion =
System.getProperty("microedition.pim.version");
```

If the optional package is not available function will return a null value. Until now the existing version is "1.0".

The first step to be able to use the PIM API is to call a static method for PIM to retrieve the singleton instance. PIM instance gives the right to access the PIM lists / databases and PIM items / fields. This is how to use them:

private PIM pim;

pim = PIM.getInstance();

The database can be manipulated after the method openPIMList() is executed. Afterthe database successfully open, the items inside can be accessed as needed. There are two kinds of methods to open the database, namely:

publicPIMListopenPIMList(intpimListType, int mode) throws PIMException

publicPIMListopenPIMList(intpimListType, int mode, String name) throws PIMException

JSR 75 can express pimListTypeas PIM.CONTACT_LIST, PIM.EVENT_LIST, and PIM.TODO-LIST. Next, for the type of access mode it can use one of PIM.READ_ONLY, PIM.READ_WRITE, or PIM.WRITE_ONLY. Open the database by name and type can be done by using the second method of openPIMList(). Items acquired after the database is opened by the following example:

ContactListcontactList = (ContactList) pim.openPIMList (PIM.CONTACT_ LIST, PIM.READ_WRITE);

Each record in the list can be accessed through the Enumeration of PIMItems. There are three methods for this, namely:

public Enumeration items () throws PIMException

public Enumeration items (PIMItemmatchingItem) throws PIMException

public Enumeration items (String matchingValue) throws PIMException

The second and third methods are commonly used for searching records based on parameters that are passed.

2.1 File Synchronization

File synchronization is a process to ensure that a file is updated each other on two or more locations with special rules. There are two kinds of file synchronization, first is one-way file synchronization and the other one is two-way file synchronization. One-way file synchronization is usually called mirroring. In this type of synchronization updated files are copied from the original site to one or more destinations, but no files to copy them back to their original locations. In contrast, the two-way file synchronization updated files are copied out in both source and destination locations, so that both locations have the exact same data.

2.2 vCard

vCard is a file format standard for storing electronic business card information. Attributes contained in the vCard generally includes name, address, phone number, email address, birthday, also with support for custom fields such as images. vCard is often used to import and export contacts in the phonebook. The PIM API also support this vCard standard. Figure 2 shows the format of vCard.

```
BEGIN:VCARD
VERSION:3.0
N:Gump;Forrest
FN:Forrest Gump
ORG:Bubba Gump Shrimp Co.
TITLE:Shrimp Man
PHOTO;VALUE=URL;TYPE=GIF:http://www.site
.com/dir_photos/my_photo.gif
TEL;TYPE=WORK,VOICE:(111) 555-1212
TEL;TYPE=HOME,VOICE:(404) 555-1212
ADR;TYPE=WORK:;;100 Waters
Edge;Baytown;LA;30314;United States of
America
LABEL;TYPE=WORK:100 Waters
Edge\nBaytown, LA 30314\nUnited States
of America
ADR;TYPE=HOME:;;42 Plantation
St.;Baytown;LA;30314;United States of
America
LABEL;TYPE=HOME:42 Plantation
St.\nBaytown, LA 30314\nUnited States of
America
EMAIL;TYPE=PREF,INTERNET:forrestgump@exa
mple.com
REV:20080424T195243Z
END:VCARD
```

Figure 2. vCard example

The data used for the application is written in csv (comma separated value) format like the above vCard example.

3. PREVIOUS WORKS

3.1 Microsoft ActiveSync

ActiveSync [5] is a protocol developed by Microsoft in 1996, allowing mobile phone users to synchronize email and calendar besides contacts. It is a proprietary standard that works with Microsoft Exchange clients which are mostly available on smart phones like Windows Phones, iPhone, Android but not on feature phones.

3.2 Google Sync

Google Sync [6], announced in 2009, uses the Microsoft ActiveSync protocol to do the same task that is synchronizing PIM (mail, contacts and calendar). It is available also only on smart phones.

The limitations of applying synchronization mechanism to feature phones can be handled by this Java ME's PIM API enabled application since it is widely available in the market for low end smart phones or feature phones.

4. IMPLEMENTATION

The programming language used in making this is Java ME applications with the NetBeans IDE 6.9 as the Integrated Development Environment (IDE). Java programming language Java ME was chosen because it is a software platform and one of the most dominant in the feature phonesecosystem today. Java ME also has cross-platform functionality and is quite safe to run on mobile devices. Besides, Java ME has an API to access personal information such as contact through PIM optional package. The use of Java ME allows creation of applications with a Graphical User Interface (GUI).

The implementation also involves a web server hosted with a public IP (Internet Protocol) address in order to store and synchronize contacts to mobiles. The web server is running an Apache with PHP support, and a MySQL database.

4.1 Synchronizing Procedure

This process consists of several sections and sub-processes. The first is to create a temporary file procedure, in which the application will create a temporary file to store contacts data that had been downloaded before the synchronization process. This procedure will try to access data on mobile phones' contacts for the synchronization process. If the access fails, then the application displays an exception that the access is denied. Another procedure is then run to continue the synchronization process. This procedure creates a csv (comma separated value) filethat is placed on the memory card slot. This file contains the contacts data that are successfully downloaded by a download from web server function.

Another procedure will open the csvfile to do the readings. Data is read per line cut and stored in an array of strings. Then the string array is passed through an add contact procedure when processing is complete for each block. This process is repeated until all data is read out and processed.

In the making of web application part, the language used was HTML, PHP, and CSS to customize the look and design. These languages were chosen because they are open source, reliable, and cross-platform. This flexibility makes it suitable for applications developed with mobile devices.

5. EVALUATION

Testing is done to see if the process of retrieving the contacts data, the contacts back-up data, and synchronization process are successful. Applications tested on four different mobile phones, the Nokia C6-00, Nokia 5730, Sony Ericsson K660i, Sony Ericsson J210i and Hazel.

Table 1 shows the success rate of installation and synchronization on Java enabled phones. Found one phone which has Java capability but the operational fails due to the JSR 75 package was not available in the phone. BlackBerry phones have Java capability but they are not using the standard PIM as in JSR 75, and application must be built separately using its own SDK (Software Development Kit). Same thing happen to Android phones.

Tests are done from moving contacts list from one phone to another, and deleting all contacts from phone then retrieve everything back from the server.

Table 1. Evaluation on mobile phones

Phones	Operating System	Java Platform	PIM – JSR 75 Availability	Success Rate
Emulator Nokia S40	S40 6th Edition FP 1	(unspecified)	ok	100%
Sony Ericsson J210i Hazel	(unspecified)	8.5	ok	100%
Sony Ericsson J108 Cedar	(unspecified)	8.5	ok	100%
Sony Ericsson K660i	(unspecified)	8	ok	100%
Sony Ericsson K810i	(unspecified)	7	ok	Operation failed
Sony Ericsson W700	(unspecified)	(unspecified)	(unspecified)	Downloading error
Sony Ericsson W8	Android 2.2	(unspecified)	(unspecified)	Have to build using different IDE
BlackBerry Curve 8520 (Gemini)	BlackBerry OS 5	(unspecified)	(unspecified)	Have to build using different IDE
BlackBerry Bold 9900	BlackBerry OS 5	(unspecified)	(unspecified)	Have to build using different IDE

5.1 Comparison to Google Sync & ActiveSync

Synchronizing with these technologies were found that existing contacts in the mobile phones are synchronized to the server and also to the devices, resulting in redundant contacts both in device and server, filling the contacts with similar entries. It seems that Google Sync and ActiveSync cannot read existing contacts or do not create a temporary table of contacts before the synchronization process.

The PIM standard on JSR 75 has an advantage over the ActiveSync on the choice of field entries on each contact. ActiveSync only synchronizes several phone numbers fields compare to the PIM which has multiple phone numbers fields as well as: address, email, birthday, photo url, and notes.

6. CONCLUSION

Based on the test results, the conclusions are:

The application is only able to retrieve the contacts list stored on the mobile phone memory not on SIM cards.

The success of the back-up data depends on the successful delivery of data and the existence and stability of Internet connections.

Application can run perfectly on the production of Nokia mobile phone with S40 6th Edition operating system and S60 (except in the Nokia E72). This may be because the application fails to access the contact list in an array of PIM database.

Application can also run perfectly on the production of Sony Ericsson mobile phones with Java Platform 8 and up.

7. REFERENCES

[1] http://www.oracle.com/technetwork/java/index-jsp-138820.html accessed 23 February 2012.

[2] Ghosh, Soma. 2009. J2ME Personal Information Management (PIM). http://www.ibm.com/developerworks/webservices/library/ws-pim/

[3] http://www.oracle.com/technetwork/java/javame/about-java-me-395899.html accessed 23 February 2012.

[4] vCard: The Electronic Business Card. 1997. Internet Mail Consortium. http://www.imc.org/pdi/vcardwhite.html

[5] DiGiacomo, Patricia. 2007. Deploying Exchange ActiveSync in Exchange Server 2007. http://technet.microsoft.com/en-us/library/bb508904(v=exchg.80).aspx

[6] http://www.google.com/mobile/sync/ accessed 23 February 2012.

Website Application Security Scanner Using Local File Inclusion and Remote File Inclusion

Agustinus Noertjahyana
Petra Christian University
Siwalankerto 121-131
Surabaya
+6231-2983456
agust@petra.ac.id

Ibnu Gunawan
Petra Christian University
Siwalankerto 121-131
Surabaya
+6231-2983456
ibnu@petra.ac.id

Deddie Tjahjono
Petra Christian University
Siwalankerto 121-131
Surabaya
+6231-2983456

ABSTRACT
Today many web-based applications developed to be accessible via the internet. The problem that often occurs is commonly found on web application vulnerabilities. Many application developers often ignore security issues when developing applications that can cause substantial losses if a hacker manages to gain access to the system. A hacker can replace web pages, obtain sensitive information, or even take over control of the website.

For that reason there is a need for applications that can help developers to overcome these problems. This application is expected to detect the vulnerabilities that exist on a website.

Existing processes include: The process of crawling to get the whole link from target websites, attacking process is useful for testing the attacks, and the last is the process of displaying a report about the security hole on the website. This application is developed using Microsoft Visual C # 2010.

Based on the results of tests made on this application, it can be concluded that the application can detect vulnerabilities in the website and report any form of link that has a security hole on the website.

Keywords
Website Application, Security Scanner, Vulnerability scanner.

1. INTRODUCTION
System security is a priority for a web administrator and web developer. But in fact web developers still prefer the design and interesting topics to increase the traffic on the website. Safety on the website is often overlooked. Though the website application security is the most important because of the presence of vulnerabilities on the website then the website will be attacked by hackers [3].

The purpose of such attacks include: extract data from websites such important credit card information / customer data, defacing the website, or even a hacker to take control of the websites server.

Based on the 2009 report from WASC (Web Application Security Consortium), an organization that conducts research in the field of web application security, attacks on the application site is increasing every year. In the report mentioned that until the end of 2009, 87% of the total existing websites still have vulnerability that can be fatal to the application site [6]. The impact of the vulnerability is not only detrimental to the developer but also detrimental to the user. Application on the website easily attacked because applications usually have different flaws and easily exploited. To help web developers in addressing these issues, it would require an application Web Application Security Scanner to detect security holes in the website and produce a report that provides a security hole on the website automatically [5].

2. WEBSITE APPLICATION

2.1 Website Application Layer
Application layer is the layer that connects a website to its users around the world [1]. At this layer there is a database containing sensitive information like credit card number, name, address, birthdate, financial records, trade secrets, medical data, and more.

At this layer is usually the main attack for hackers who have a variety of purposes, such as information theft, damage the website, or even take over control of the website. This layer was attacked because it is the weakest layer than other layers. The rapid development of technology can make more complex the application layer. There will be lots of bugs and the possibility of security holes that can be exploited by hackers to get into the system.

From the report WASC (Web Application Security Consortium), said that currently 75% of cyber attacks in the world begins at the application layer and can adversely impact on the website [6].

2.1.1 Website Application Security
Web application security is a way to protect the site from the application layer attacks that can cause losses to a variety of individual and corporate [7].

2.1.2 Website Application Security Scanner
Web Application Security Scanner is the software that automatically searches for vulnerabilities that exist on the website. This software does not access the source code in carrying out its action, but using the Black Box [5].

Black Box is a step in which the application will perform experiments to attack and then capture the response of the attempted attacks have been carried out, after which the application will process and evaluate the response to a report in the form of an overview of security loopholes that exist on the website.

This method is commonly known as Penetration Testing.

These applications are generally divided into three stages in carrying out its action, namely Crawling Component, Component Attack, and Analysis Modules [5].

- Crawling Component, which is popular with the term Web Crawler, the website will index the links on the website. This step can be performed by various methods, depending on the needs of application users.
- Attack Component, in which the application will start automatic attempted attacks on a link that has been indexed.
- Analysis Modules in which the application will evaluate the responses provided by the website and create reports of website security picture.

2.2 Web Crawler

Web crawlers are also commonly known as the Spider Web. This method collects all the information on web pages. Web Crawler log every link on the website pages visited, and then browse through these links one by one. In its implementation, there are various methods of web crawlers that are used as required [7].

The methods commonly used on the Web Crawler are:

- BFS (Breadth First Search)
  Search based on the breadth of website information, on the implementation uses queue as the storage of the URL
- DFS (Depth First Search)
  Search depth information on the website, on the implementation uses stack as the storage of the URL

A Web crawler process generally begins with providing the initial set of URLs as the beginning of a search into a queue. Priority of the criteria can be applied to reorder the list of URLs in the queue. Furthermore Crawler downloads pages by URL from the queue. Once deposited into the collection, obtained parsed pages to get the out-going link and subsequently inserted into the queue. The process continues until the queue is empty or if the stop condition has been fulfilled. Figure 1 shows a generic web crawler algorithm.

```
Input: Seed = {u1, u2, ..., un} daftar URL awal
URL_Pool ← Seed
Visited ← ∅,  URL yang telah di kunjungi
while URL_Pool ≠ ∅
    u ← Select (URL_Pool, Kriteria Pemilihan)
    p ← Download (u)
    Visitied ← Visited ∪ u
    out_link ← Extract_Outgoing_Link (p)
    for each q ∈ out_link
        if (q ∉ Visited) and (q ∉ URL_Pool)
            URL_Pool ← URL_Pool ∪ q
        end if
    end for
end while
```

Figure 1. Generic Web Crawler Algorithm

2.3 File Inclusion

File Inclusion is a method for attacker to insert malicious code into a site that has security holes, where penetration is done by two ways:

- Local File Inclusion (LFI)
- Remote File Inclusion (RFI)

2.3.1 LFI (Local File Inclusion)

Local File Inclusion is a web application flaws that led to an attacker could exploit the Web Server by accessing the directory and execute the command in web server other than root. [2].

Local File Inclusion closely related to the Directory Traversal. LFI performed locally from the web server. LFI normally occurs due to the variables during of website creation is poorly controlled thus allowing an attacker to access the folder directly.

Functions which could lead to Local File Inclusion (LFI) on the website are:

- include ();
- include_once ();
- require ();
- require_once ();

With the condition in the php configuration on the server:

- allow_url_include = on
- allow_url_fopen = on
- magic_quotes_gpc = off

Example code:

```
<? php
$ page = $ _GET ['page'];
include ($ page);
?>
```

If the function is accessed by the LFI methods of:

- http://victim.com/index.php?page=../../../../../../../etc/passwd

It will display the contents of / etc / passwd, where '.. /' is a method to move one folder up from the current folder. Development of the LFI method is Remote Connect-Back Shell, where an attacker gains a shell of the target server and the attacker could perform a variety of linux commands on the existing system.

2.3.2 RFI (Remote File Inclusion)

Remote File Inclusion is the type of attack that allows an attacker to insert file from outside of server. These attacks have a harmful impact because the file is inserted in the form of shell [2]. By utilizing this Shell a hacker can gain access to the system that perform various activities such as viewing directory, change the page (defacing) or steal sensitive information from existing files in the system. Here's an example of writing in the RFI website:

- http://deddie.com/page.php?page=http://site.attacker/kode.txt

If the site does not have good control in the the variables $ page, then the contents of “kode.txt” will be displayed in the browser. The contents of the kode.txt can be modified by an attacker to install the backdoor shell on the target system.

Typically the attacker will try to use the Remote File Inclusion first to penetrate in the website, but if the attacker knew that a form of penetration that he did fail (usually due to “allow_url_include = off” in the php.ini), will pursue an attack by

a method in which the page Local File Inclusion inserted on the same server.

File Inclusion form of attack is high risk, because the attacker could gain shell access, and ultimately affect Local Exploitation in which the attacker gains full access rights to the system.

3. DESIGN

The design of this scanner security system applications are: starting the process of crawling to get a website structure, conduct experiments attacks on the pages that have security holes, and displays the results of security holes reports from the website.

The design of the application security scanner can be seen in figure 2.

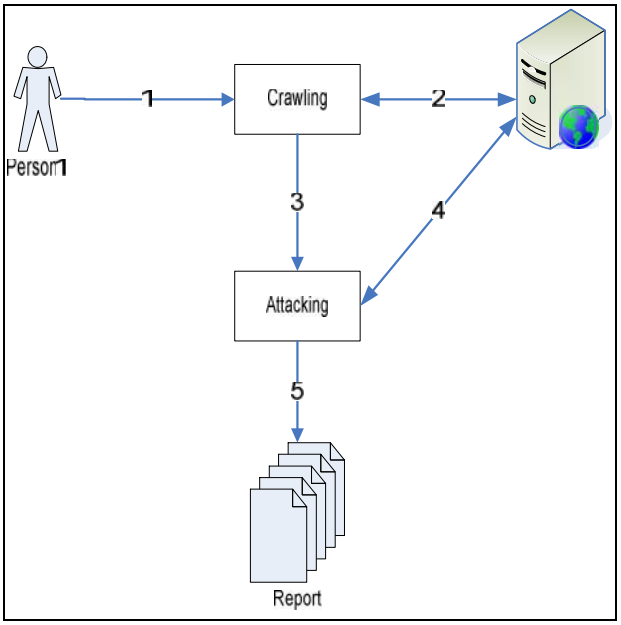


Figure 2. Design of thr Application Security Scanner

3.1 Crawling Process

Crawling process is a process where the software scans the response from the server to get the links on the website. process is performed recursively, allowing the software to index the entire link from the website. The system provides additional features that allow users to see the crawling process, log, and the structure of the crawled website, and see a link that is ignored. Crawling process can be seen in Figure 3.

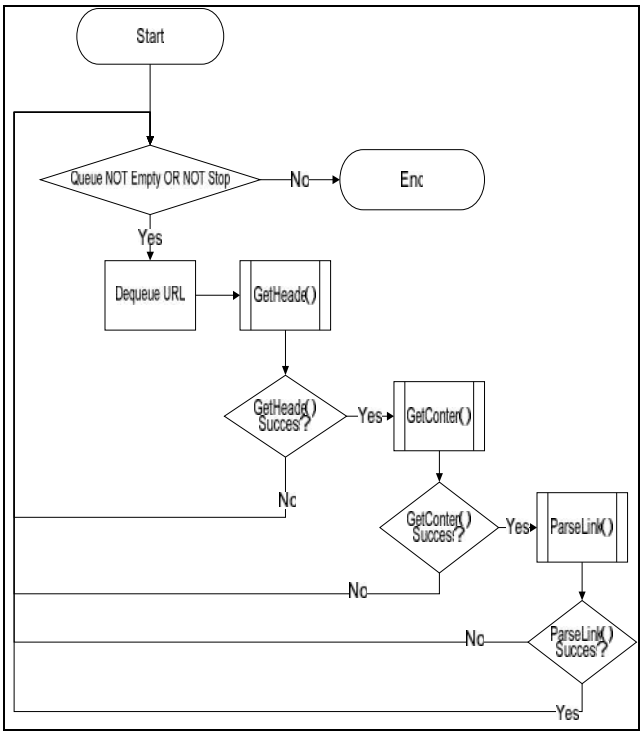


Figure 3. Crawling Process Diagram

3.2 Enqueue URL

Enqueue URL process used in the crawling process is not obtained directly insert a link to the Queue. The initial steps of this process are performed by checking whether the protocol was obtained from a valid link. This is done because there are different styles of writing on any web developer to create a website. Differences in writing style makes the software must create an additional process that links can be changed to a valid link before it is inserted into the Queue.

In the Enqueue URL process also checking whether the link is still to be obtained in the same website. It aims to restrict the software to be in the process of crawling, do not crawl links from different websites. Flowchart of this process can be seen in Figure 4.

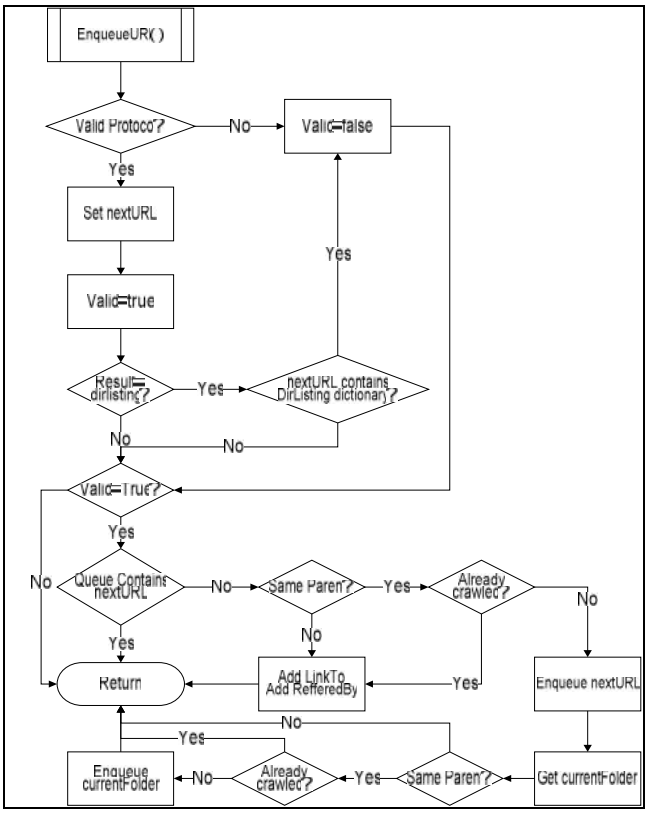


Figure 4. Flowchart of Enqueue URL Process

3.3 Attacking using LFI & RFI

Attacking Process is a process in which the experiments will be conducted attacks on the link that has been obtained from the crawling process. The result of this process is to detect whether the link has a security holes, and then will make a report about the security holes of the website.

In this process, the software will try to attack using the LFI and RFI method. Flowchart to show the LFI and RFI can be seen in Figure 5.

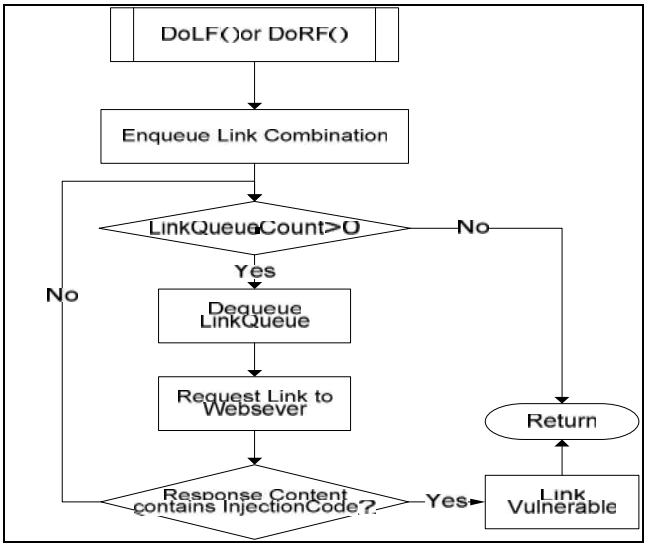


Figure 5. Flowchart of LFI and RFI method

3.4 IMPLEMENTATION

In this section will describe the implementation of application security scanner.

The first time when running the application, the user is asked to perform global configuration settings as shown in figure 6.

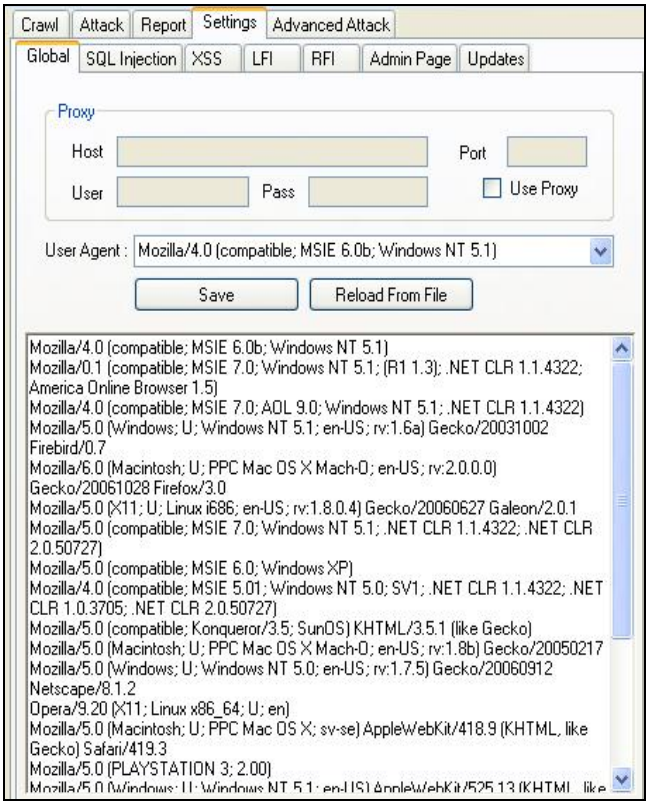


Figure 6. Global Configuration

The main process starts from the crawling process on the website that has been conditioned to have security holes. The first step is to get all the links that exist in the target website by crawling process, which can be seen in Figure 7.

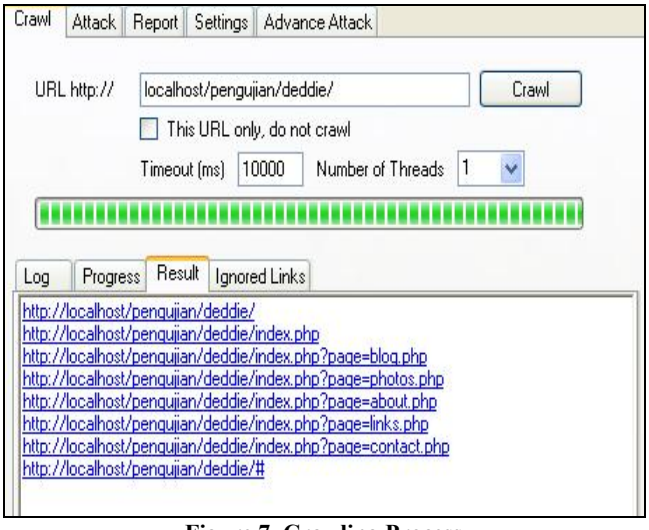


Figure 7. Crawling Process

The next process is to perform database modification of the LFI and RFI database. LFI database modifications can be seen in figure 8, while the RFI database modifications can be seen in Figure 9.

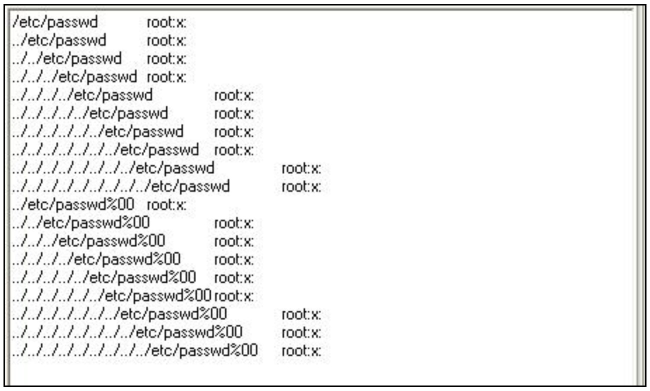


Figure 8. LFI database modifications

Figure 9. RFI database modifications

In the attacking process, testing is done by changing the configuration of LFI column for testing using localhost. Then the application will try to open a file passwords.txt located in XAMPP folder by using the string root as a marker.

In the RFI testing, trying to open the address "http://google.com" using "Google <title> <title>" string, as a marker of security holes. Configuration is shown in figure 10.

Figure 10. Configuration of LFI and RFI test

The results of the attack can be seen in Figure 11.

Figure 11. Result of Attacking Process

In the figure 10 shows that there are security holes in the LFI and RFI. For those reasons, the testing will proceed with trying to manually request the link. In this test uses link "http://localhost/pengujian/deddie/?page=../../../../xampp/passwords.txt." to prove the security holes that exist. The results of this test can be seen in Figure 12.

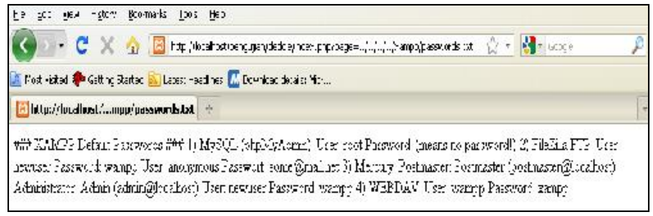


Figure 12. Result of LFI Test

RFI test by doing manually in the browser request to prove the validity of the application. Link that selected for testing was "http://localhost/pengujian/deddie/?page=http://google.com". The results of the RFI test can be seen in Figure 13.

Figure 13. Result of RFI Test

4. CONCLUSION

- The crawling process can affect the outcome of the report is generated; this is because the link that will be processed on the next stage (Attacking) is a link that has been obtained in the process of crawling.
- In the field testing LFI vulnerability allows an attacker to open any file by simply entering the URL path. When the website was found loopholes in this section attacker could potentially get the contents of sensitive files, which usually are targeted is a file containing the configuration of the web server.
- In testing the security gap in the field of RFI, can be seen that on a website that has this vulnerability could inject a file from another website. This would allow an attacker to inject php shell to the target website. Where the php shell, the attacker can mess up the contents of the website, it can even take over control of the target website.

5. REFERENCES

[1] Black, Paul et al. 2008. Software assurance tools: Web application security scanner functional specification version 1.0. National Institute of Standards and Technology. United States.

[2] Heydari, Mir Seyed Ali. 2008. A new source code auditing algorithm for detecting LFI and RFI in PHP programs. World Academy of Science, Engineering and Technology. Izlamic Azad University, Iran.

[3] Huang, Yao-Wen et al. 2004. Non-detrimental web application security scanning, ISSRE, pp.219-230, 15th

International Symposium on Software Reliability Engineering (ISSRE'04).Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[4] Johns, Martin et al. 2008. XSSDS: Server-side detection of cross scripting attacks, 24th Annual Computer Security Applications Conference (ACSAC '08), pp. 335 - 344, IEEE Computer Society.

[5] Kals, Stefan, et al. 2006. Secubat: A web vulnerability scanner. Technical Security University of Vienna.

[6] Orloff, J. 2009. Web application security: Testing for vulnerabilities. New York: Sequoia Media Services. Inc.

[7] Stasiak, K. 2002. Web application security. Ohio: SecureState, Inc.

The Development of Web Security Scanner Based on XSS and SQL Injection Method

Ibnu Gunawan
Petra Christian University
Siwalan kerto 121 - 131
Surabaya
+62312983456
ibnu@petra.ac.id

Agustinus Noertjahyana
Petra Christian University
Siwalan kerto 121 - 131
Surabaya
+62318439040
agust@petra.ac.id

Deddie Tjahjono
Petra Christian University
Siwalan kerto 121 - 131
Surabaya
+62318439040
m26407051@john.petra.ac.id

ABSTRACT
Nowaday, there is so many vulnerabilities in web application layer. This is because of security issues that are often overlooked by a web developer when creating a website. In fact, caused by the presence of vulnerabilities on a website, a hacker can do a variety of activities that destroy of website. Adverse events that can be done by a hacker includes changing the web page (defacing), obtain sensitive information, even taking over control of the website system. To help overcome these problems, we make an application to detect vulnerabilities that exist on a website.

The process is started by crawling to get the entire link from the target website. Followed by attacking the process that is useful to attempt an attack on a link that has the potential security hole. The application will then continue in the process of reporting where the application would create a vulnerability report on the website. This application was built using Microsoft Visual C # 2010.

Based on the results of tests made on this application, it can be concluded that the application can detect vulnerabilities in the website and report any form of link that has a security hole on the website.

Keywords
Web, application, security, scanner, xss, sql, injection

1. INTRODUCTION
Security of the application site should be a priority for a web administrator and web developer. But generally the manufacturer's website only give priority to the design and what topics are provided in order to attract as many visitors. Website security is usually placed on the nth order. Though the website application security is the most important because of the presence of vulnerabilities on the website then the website will be attacked by hackers [1].

Based on the 2009 report from WASC (Web Application Security Consortium), an organization that examines the field of web application security, attacks on the application site is increasing every year. Which, from his report said that until the end of the epidemic in 2009, 87% of the total existing websites still have gaps that can be fatal to the application site [2]. Of the report, said that the security hole is in the most XSS (Cross-Site Scripting) as much as 39%, followed by 32% Information Leakage, SQL Injection and as much as 7%.

The impact of the gap is not only detrimental to the developer but also detrimental to the user. The number of attacks in this layer due to the application site is very easy to attack because in general the application site has many weaknesses and easily exploited.

Therefore, to help web developers in tackling this problem, we need Web Application Security Scanner for detecting a variety of security holes in the website and produce a report in the form of a report containing an overview of the security holes in the website automatically [3].

2. PRELIMINARIES
In this section we briefly introduce about web application layer, its security, and web application security scanner. Then we moved to it's method especially the sql injection and xss.

2.1 Website Application Layer
Website or application layer can also be called the application layer is a layer of websites that act as primary liaison with the website users around the world [4]. There is also a database that contains highly sensitive information like credit card number, name, address, date of birth, *financial* records, trade secrets, medical records, and many other important information.

This layer is usually the main attack for hackers who have a variety of purposes, such as information theft, damage the website, or even take over control of the website. This layer was attacked because it is the weakest layer than other layers. Along with the development of technology, the automatic will indirectly create a more complex application layer. Where this is going to make so many of them the possibility of bugs and loopholes that can be exploited by hackers as a way into the system.

From the report WASC (Web Application Security Consortium), said that currently 75% of cyber attacks in the world begins at the application layer and the website can be fatal to the website [2].

2.2 Website Application Security
Web application security is a range of measures taken to protect the application layer on the website of the hacker attacks that can cause a variety of losses for individuals and companies the website owner[5].

2.3 Website Application Security Scanner
Web Application Security Scanner is the software that automatically search for vulnerabilities that exist on the website. This software does not access the source code in carrying out its action, the means used this software to detect security loopholes

that exist is by Black Box [3]. Another name for this kind of application is a Web Application Vulnerability Scanner.

These applications are generally divided into three stages in carrying out its action, namely Crawling Component, Component Attack, and Analysis Modules [3]:

- Crawling Component or a term popular with the Web Crawler, where the website will index the links on the website. This step can be performed by various methods, depending on the needs of application users.
- Attack Component, in which the application will start automatically attempted assault on a link that has been indexed.
- Analysis Modules in which the application will evaluate the responses provided by the website and create reports of website security picture

2.4 Web Crawler

Web crawlers are also commonly known as the Spider Web. Where this method has a duty to collect all the information in our website. Work performed by the Web Crawler is done automatically by the record of each link on the website pages you visit and then visit these links one by one. Its implementation, there are various methods of web crawlers that are used as needed [6].

The methods commonly used in Web Crawler is:

- BFS (Breadth First Search)
  Search based on the breadth of available information website, which on its use as a storage utilizing URL Queue
- DFS (Depth First Search)
  Search depth of information available on the website, which on its use as a storage URL Stack harness

In short, a Web crawler process generally begins by providing a set of initial seed URL as the search into a queue. Priority criteria can be applied to reorder the list of URLs in the queue. The next crawler to download web pages by URL is retrieved from the queue. Once deposited into the collection, obtained parsed pages (parsed) to be extracted out-going unvisited link and then inserted into the queue. Pick-up process continues until the web page URL queue is empty or if the stop condition is met. Figure 2.1. shows a web crawler.

Input: $Seed = \{u_1, u_2, \ldots, u_n\}$ daftar URL awal
$URL\_Pool \leftarrow Seed$
$Visited \leftarrow \emptyset$, URL yang telah di kunjungi
while $URL\_Pool \neq \emptyset$
  $u \leftarrow$ **Select** (URL_Pool, Kriteria Pemilihan)
  $p \leftarrow$ **Download** (u)
  $Visitied \leftarrow Visited \cup u$
  $out\_link \leftarrow$ **Extract_Outgoing_Link** (p)
  for each $q \in out\_link$
    if $(q \notin Visited)$ and $(q \notin URL\_Pool)$
      $URL\_Pool \leftarrow URL\_Pool \cup q$
    end if
  end for
end while

Figure 2.1 Web Crawler Genetic Algorithm

2.5 Regular Expression

Regular Expression or more often called Regex is a technique used to match a text string, such as particular characters, words, or patterns of characters. Regex has two main functions, ie search and replace, find a certain pattern in the text and then switch to another patter [8]. Some common patterns used in the regex shown in table 1:

Table 1. Common pattern used in the regex

Symbol	Meaning
*	Replace character to infinte
+	Replace one character to infinte
?	Replace character 1 and 0
^	Search for a word that begins by pattern
$	Search for a word that ends the pattern
|	Give a choice
()	Make sub pattern

2.6 SQL Injection

SQL Injection is a technique that utilizes SQL query writing errors on a website so that a hacker could add some SQL statements to the 'query' by manipulating data input into the application. So that the database server to generate an invalid SQL query[7]. On the reality, SQL Injection is a proven one of the best techniques that often paralyze the target. With this technique the attacker can log into the system without having an account. Figure 2.2 show a SQL Injection syntax

```
http://10.252.108.232/web1/index.php?option=product.php&status
=1;update barang set harga = 50 where barangID=9;
```

Figure 2.2 SQL injection via URL.

2.7 XSS (Cross Site Scripting)

Cross Site Scripting is a type of attack where the method used is to inject Javascript into a website. Attacks of this type is usually underestimated because in most cases have an impact on the client or the so-called Client-Side Script. But in fact this kind of fatal attack[9], because an attacker could potentially do the following:

- Users can inadvertently run a script that has been inserted by the attacker and open the content according to the script.
- The attacker can take over from the user's active session before the session expired. Where the impact is the attacker can get into user accounts without having to make the login process.
- Attackers can connect users automatically to the server designated by the attacker.

to know that the website has a XSS vulnerability we can use method of Request / Response Match[10]. Where this method is trying to insert XSS code in the URL and make requests to the webserver. When the webserver responds in the form of content that contains XSS code, it can be said that the website has a security gap in this field. Figure 2.3 show a flowchart of Request / Response Match method

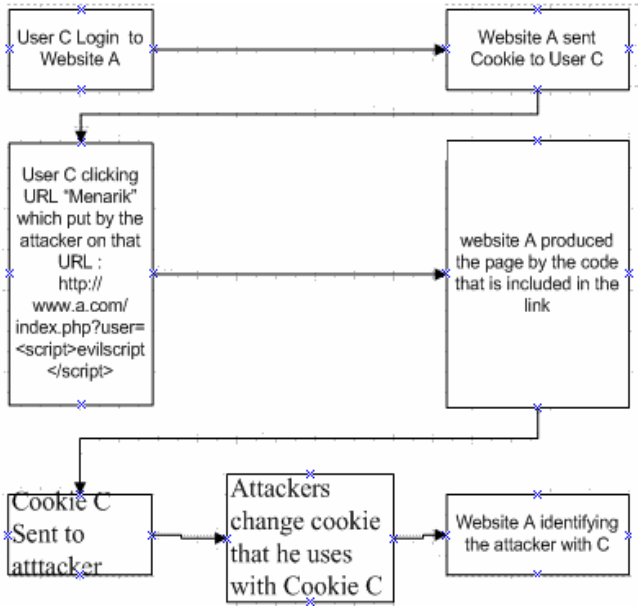


Figure 2.3 Flowchart of request / response match method

If the XSS injection performed on a variable that simply pass a parameter without saving it in the database, then the results are only temporary (temporary). But if this weakness is found in the Guest Book, Shout Book, Forum, Blog, and the like and do XSS attack, the result of such attacks would be permanent because the injected script is stored in the database.

3. SYSTEM DESIGN

The software has been developed is consists of several stages. Start by crawling process to get a website structure, conduct attacks experiments on the pages that have the potential to have security holes, and displays the results of an overview report of website vulnerabilities. Figure 3.1 show the main flowchart.

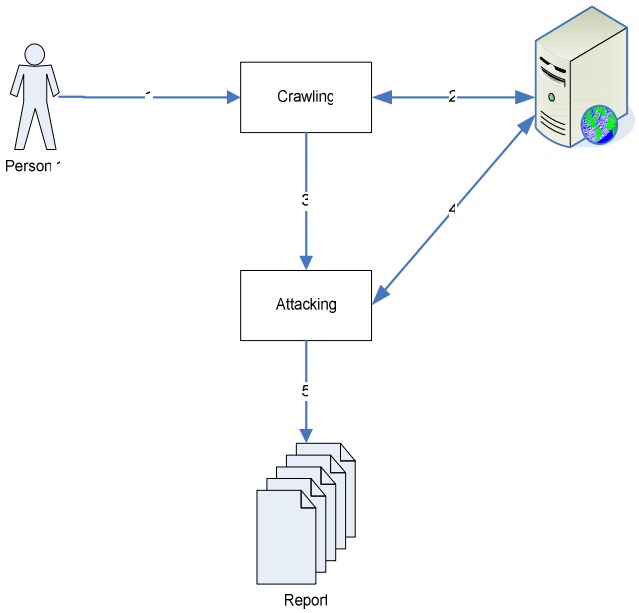


Figure 3.1 Main system flowchart

3.1 Threading Process

At the beginning of each function to be executed, held the settings thread, where the software does is to ask the user how many threads and timeout to be used in the process. Timeout setting is used as the reference length of the website provides a response. While the threads are useful to accelerate the setting process of a function. Figure 3.2 show the main threading process

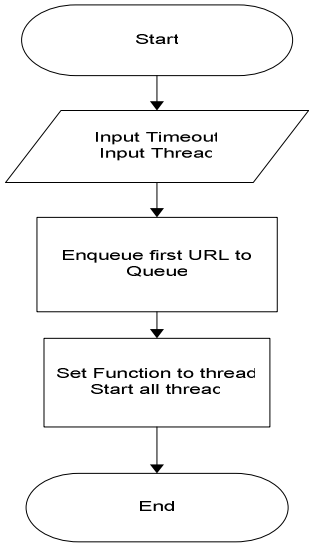


Figure 3.2 Main threading process

3.2 Crawling process

Crawling process is a process where the software will scan the results to the response given from the server to get the links on the website. Where the process is done recursively, allowing the software to index the entire link from the website. The system provides additional features that allow users to see the running processes Crawling, saw logs and structure of the website is being crawled, and see the link is ignored. Figure 3.3 show the main flowchart for the crawling process.

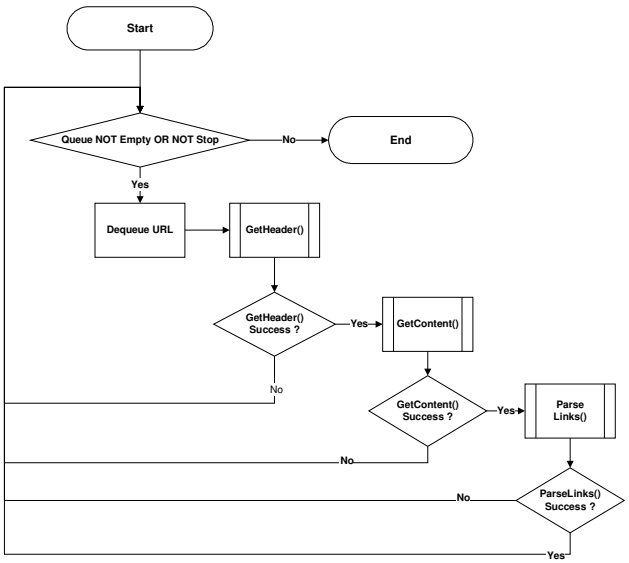


Figure 3.3 Main flowchart crawling process

3.3 Attacking Proccess

Attacking process is a process in which the experiments will be carried out attacks on the link that has been obtained from the crawling process. From the results of such an attack would be detected if the link is being attacked have security holes in certain areas. The results of this process will be sent to the Report to create reports on website security picture.

At this process, the software will ask the user what type of experiments that will be attempted attack on the target website. Where the trial will be conducted in accordance with user's choice. Figure 3.4 show the main process of attacking process

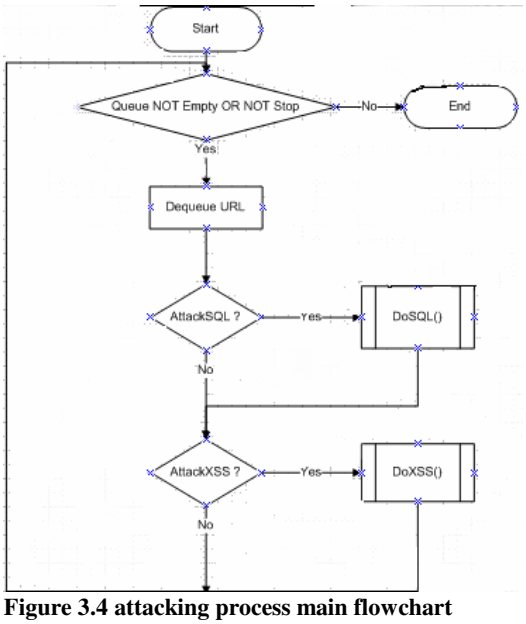


Figure 3.4 attacking process main flowchart

4. IMPLEMENTATION

Web application is made using the programming language C # using Microsoft Visual C # 2010 as an Integrated Development Environment (IDE) from the C # language. The reason the use of the language C # is because the ability of C # which is known to OOP (Object Oriented Programming) is fairly easy to apply and the language is quite easy to implement, as well as the compiler Microsoft Visual C# are very supportive of designing GUI (Graphical User Interface) and very flexible to use.

Whole process is made using a namespace that already provided by Microsoft Visual C# 2010. This is because C # does not have its own class library, so the use of class library in C # using a class library that is used in Visual Basic and Visual C + +. The namespace is used as follows:

- System.Threading, used for setting the thread that will be used to accelerate the course of these functions in the application.
- System.Text.RegularExpressions, used to set the regular expression pattern that will be used to detect the link on the website content
- System.IO, used for setting the Save and Load data from files that are used in applications. Including storage of the final report as a result of the application.
- System.Net, is used to request a link to the web server settings

Figure 4.1 Show the class structure that has been made for this web application.

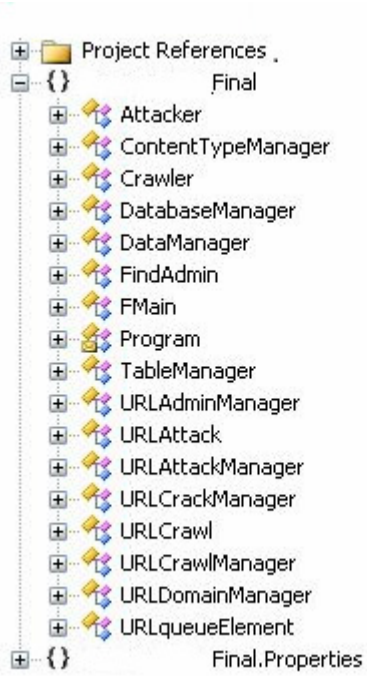


Figure 4.1 Class structure diagram

The main classes used in the application is:

- Crawler, a class that contains functions to perform the process of crawling on the target website.
- URLCrawlManager, a class that contains functions for setting the course of the process of crawling, where the class is setting the thread to do the crawling.
- URLCrawl, a class that is used to store all data generated by the crawling process

- URLQueueElement, a class that contains the data to be inserted into the queue used to process the application.
- The attacker, a class that contains functions to perform the process of attacking the target website.
- URLAttackManager, a class that contains functions for setting the course of the process of attacking, in which the class is setting the thread to do the attacking.
- URLAttack, a class that is used to store all data generated by the process of attacking
- URLDomainManager, a class that is used to store the domain obtained during the application process.
- URLContentTypeManager, a class that is used to store the content type of link that obtained during the process of crawling on the application.
- FindAdmin, a class that contains functions to perform the search process on the target website admin page.
- URLAdminManager, a class that contains functions for setting the course of the search process admin page, where the class is setting the thread to the admin page to do searches.
- URLCrackManager, a class that contains functions to perform the process Crack Data on the target website.
- DatabaseManager, a class whose function is to store the entire database is available on the Crack Data. TableManager, a class whose function is to store the entire database structure is successfully obtained.
- DataManager, a class whose function is to store all data from a database which is found in the Crack Data

5. PROGRAM TESTING

A first step the use of a user's system will automatically be transferred to the tab Crawl, where the tab is a link the user to enter input or website name to be researched vulnerabilities. The initial steps to be performed is the process of crawling. Crawling is the process of the initial stage where the application will try to make repeated requests to get all the links on the website. Users can also set the number of threads used options at the time of the Crawling.

Figure 5.1 show menu of the main screen of our application, figure 5.2 show the option for proxy and user agent, figure 5.3 show the crawling process. Then we continuing our attack using xss which is setting is shown in figure 5.4 after that we continuing testing using sql injection which is setting is shown in figure 5.5. the result is shown in figure 5.6, 5.7 and 5.8 for reporting.

Based on experiment that have been done, we can see that xss and sql injection can caused enough trouble for non hacker prepared site.

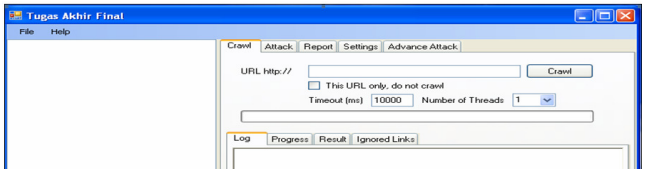


Figure 5.1 Web App Vulnerability scanner main screen

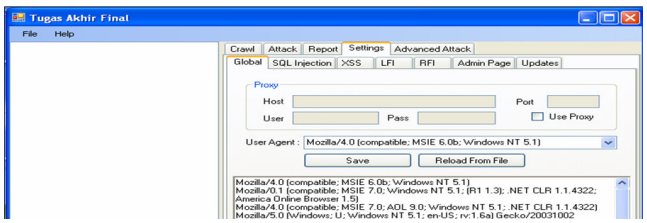


Figure 5.2 The option for proxy and user agent

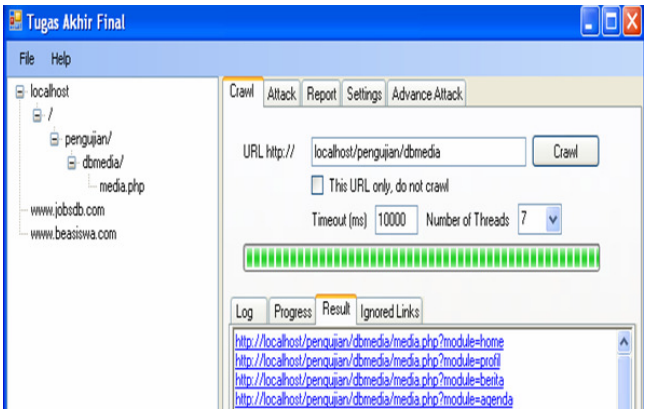


Figure 5.3 Crawling process

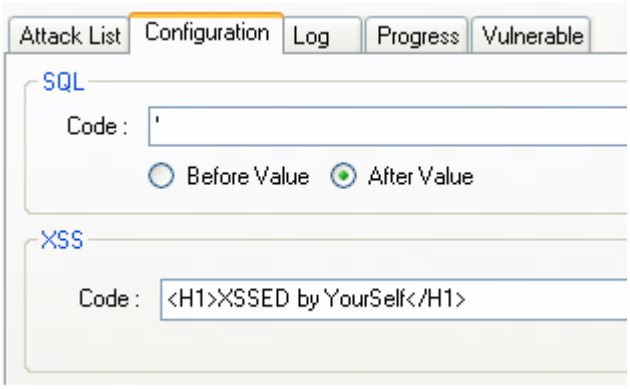


Figure 5.4 XSS setting

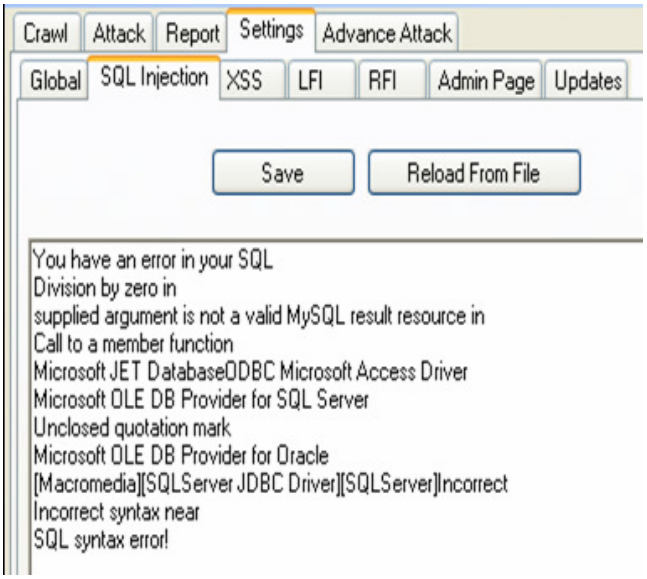


Figure 5.5 SQL injection setting.

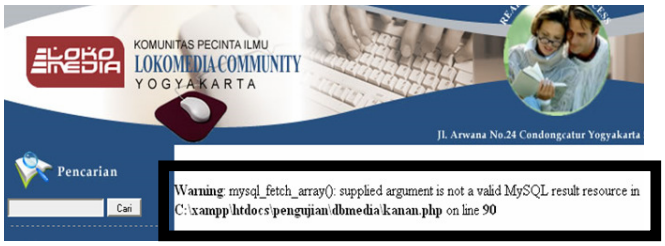


Figure 5.6 sql injection result

Figure 5.7 XSS attack result

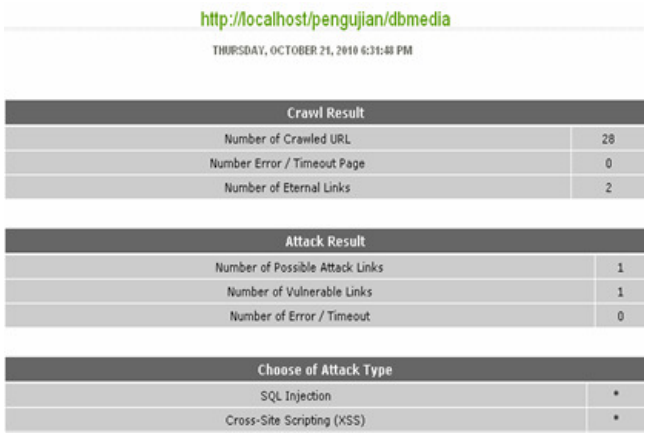


Figure 5.8 Attack result report

6. CONCLUSION

Based on systems that have been developed and test results that have been done, we can conclude some of the following:

- The crawling process can affect the outcome of the report is generated, this is because the link that will be processed on the next stage (Attacking) is a link that has been obtained in the process of crawling
- The test can be seen that application of the flaw could be used as link testing. It has been proved by testing manually obtained from the link.
- The first test to prove the security hole in the areas of SQL Injection can be exploited by attackers to get the entire database structure and data-sensitive data from target website.
- In the first test also seen the gap XSS flaws, where the gap is an attacker can inject javascript. It can be dangerous because an attacker can modify it for various uses javascript attacks.

7. REFERENCES

[1] Huang, Yao-Wen et al. 2004. *Non-detrimental web application security scanning,* ISSRE, pp.219-230, 15th International Symposium on Software Reliability Engineering (ISSRE'04).

[2] Orloff, J. 2009. *Web application security: Testing for vulnerabilities.* New York: Sequoia Media Services. Inc.

[3] Kals, Stefan, et al. 2006. *Secubat: A web vulnerability scanner.* Technical Security University of Vienna

[4] Black, Paul et al. 2008. *Software assurance tools:Web application security scanner functional specification version 1.0.* National Insitute of Standards and Technology. United States.

[5] Stasiak, K. 2002. *Web application security.* Ohio: SecureState, Inc.

[6] Widiantoro, Dwi, H. 2006. *Survey arah penelitian, pengembangan, dan penerapan penjelajah situs web.* Proceeding of International Conference on Instrumentation, Communication and Information Technology, Insitut Teknologi Bandung.

[7] Sunyoto, A. 2004. Metode penyerangan website menggunakan SQL Injection.*Jurnal DASI* 5 (3), Retrieved 8 November 2010 from http://journal.amikom.ac.id/index.php/informatika/article/viewArticle/124

[8] Cho, J. 2002. *A fast regular expression indexing engine.* Proceeding of 18th International Conference on Data Engineering. University of California.

[9] Saha, S. 2009. *Consideration points: Detecting cross site scripting. International Journal of Computer Science and Computer Security (IJCSIS)*, Hanyang University.

[10] Johns, Martin et al. 2008. *XSSDS: Server-side detection of cross scripting attacks*, 24th Annual Computer Security Applications Conference (ACSAC '08), pp. 335 - 344, IEEE Computer Society.

Optimization of Scheduling Based on Tasks Merging Technique

Marjan Abdeyazdan
Department of Computer Engineering, Mahshahr branch, Islamic Azad University
Mahshahr, Iran
Phone: 0098-916-152-0433
E-mail: abdeyazdan87@yahoo.com

ABSTRACT
The issue of static scheduling of tasks is of particular importance in multiprocessor systems due to optimal use of processors and taking less time. It is a NP-Hard problem and obtaining the optimal solution involves a high time complexity. Thus the heuristic method is used to solve this problem. The genetic algorithms and Automata are suitable method for scheduling of multiprocessor systems. In this paper a solution is introduced for algorithms of scheduling. The main idea in designing these algorithms is to obtain the least run time and the highest possible parallel with minimum critical path. By using of this solution the graphs with many tasks can be transformed in to fewer ones. We use a combination of both tasks merging technique and scheduling with genetic of algorithm to shorten the length of critical path.

A new genetic algorithm is introduced to schedule of multiprocessor systems in which the scheduling priority of performing the tasks is based on the number of child. The results show that the new proposed algorithm achieves the optimal scheduling solution in acceptable time compared to other common methods.

Keywords
Scheduling, tasks graph, merging, multiprocessors, genetic algorithm.

1. INTRODUCTION
The issue of scheduling in multiprocessor systems and heterogeneity distributed systems is a NP –Hard problem. In the classical method it is time-consuming to obtain of a complete optimal solution of a minimum scheduling and in many cases running the tasks randomly takes too much time. Thus the heuristic complete optimal solution is not necessarily obtained but within reasonable time span an answer will be obtained which almost close to optimal solution. Many heuristic methods have been introduced include: Duplex[14,12], OLB[14,12,4], GA[19,20,17,24,21], SA[1,2,19,13,3], LMT[11], MCT[14], A*[16,2,19,15], Min-min[14,12,10], Max-min[14,12,10], DLS[5], GSA[13], Tabu search[18,19]. One of the best heuristic method is to use the genetic algorithm. In this paper a new genetic algorithm based on earliest tasks run is simulated given their priority on the number of child.

In order to increase the program efficiencies in parallel systems an accurate scheduling of tasks is highly taken into consideration. The scheduling should be performed so that the total run time of the program with regard to tasks run time and the relation between processors can be minimized. Certainly the more the number of processors during running, the simpler the planning carries out. Today it is common to use a lot of processors in super computers, quick processing clusters, the distributed networks like grids. Hence it is important to optimize along the critical path. By using the tasks merging the length of this path can be decreased considerably.

Given being NP-Complete of the scheduling task the approaches based on uncertainty methods in this field will not ever efficient. In the past decades much works have been performed by this algorithm, however many of which had not the required efficiency. Therefore, use of uncertainly method and mainly the genetic algorithm is very effective for solving such problems.

In this paper, we use a combination of both task merging technique and scheduling in order to shorten the critical path. Accurately merging of the tasks before scheduling the task graphs have significant impact on scheduling result. The results obtained by applying our algorithm on the different graphs for evaluating the scheduling algorithm represent a positive impact of merging on scheduling.

2. RELATED WORKS
In order to improve the performance of scheduling algorithms of task graphs, before applying the algorithms, they can be prepared by using two solutions i.e. clustering [6]and tasks merging [22,26] for suitable scheduling. The results from multi processors scheduling algorithms can be greatly improved by using of both solutions. In the following, the two above solutions will be examined to solve multiprocessors scheduling problems.
A lot of works have been done on task scheduling in the multiprocessor systems by genetic algorithm [19, 20, 17]. Some of them have employed the random genetic algorithm and some others used genetic algorithm and prioritizing tasks based on height [20]. In this research a new genetic algorithm has been simulated based on earliest task run given their priority and also on the number of their child.

2.1. Clustering the Tasks
A task clustering solves only a part of multiprocessor scheduling problem. A cluster of tasks is a set of tasks that together are run on a processor. In turn all the edges inside a cluster of a graph their relation cost equals to zero. Although the sequence of running tasks within a cluster is often determined by algorithms of clustering, the main aim of them is to increase granulating of tasks graph. However the tasks within clusters are not summed together. In the other word, the resulting data from each of tasks are sent as soon as provided. It is considered as a substantial difference between task clustering technique and task merging [2].

In order to cluster the tasks some optimal or nearly optimal processors are determine for task graph scheduler. In the other word, the more processors are needed from generated clusters by algorithm of task scheduling. Also it shows where the task clustering is performed prior to scheduling. The balancing (LB) for merged cluster is proffered because LB is fast and is easily implemented and yields better final scheduling. In short, two phase method of jobs scheduling and balancing have significant effect strategically for job graph scheduling on distributed parallel architecture.

2.2. Algorithm of Merging Task

In [4] a merging task algorithm is presented which its input is a graph with micro granulating. The main idea is iterative selection of a pair node for merging. If merging of these two nodes lead to decrease the length of critical path, then it select two new nodes.

2.3. Algorithm Based on Graph Rewriting System

Graph Rewriting System (GRS) Is an algorithm used for merging tasks [22]. It is employed as a way for reducing the length of critical path. The transformation is specified by several graph rewriting rules.

2.4. Prioritizing Scheduling Algorithm Based on Height or Based on Offspring (Children)

The aim of scheduling in a multiprocessors system is to assign N job to M processor so that the finished run time of final job in this system is minimized. For simplicity, if two tasks are scheduled on two different processors, then the communication cost for sending data between two tasks is taken as zero. The algorithm of task scheduling generation in term of height or offspring is as follows:

a. Arrange the tasks in order of incremental in term of their height or offspring (children).

b. Repeat the stages 3,4 in the tasks.

c. Generate a random number r, so that 1<r<m (m is the number of processors)

d. Select the first task from arranged ones and allocate it to processor r^{th} then eliminate it.

By repeating the algorithm of scheduling generated, an initial population is generated from search nodes. The path from input node to output node is the longest critical path where we denote it by and length of path by representing the run time of the program in a state in which the maximum number of processors can be used. (Each existing task in task graph is sent to a processor) the value of upper bound of output node reflects the length of critical path of the graph.

3. PROPOSED ALGORITHM

This novel scheduling algorithm is a combined technique resulting from task merging as well as tasks scheduling. In order to manage large graph with micro granulating, the tasks are merged so that in addition to improve the graph specifications such as length of critical path, the amount of granulating is increased as well. It is done through graph rewriting system which consist the series of codes transforming graph on task graph.

It is NP-Complete. Therefore, the approaches based on conclusive methods are not so efficient in this field. Using of evolutionary processing algorithm and mainly genetic algorithm given their non-conclusive nature will be effective for solving problems. Genetic scheduling algorithm is ever seeking to find the closest answer. In order to find a closest answer, this paper presents a two-step algorithm as follows:

The first step is to receive a task graph with micro granulating. After receiving the task merging algorithm graph, it analyzes the graph and attempts to merge the tasks for modifying the parameters of qualitative program. The next one, the obtained graph from this merging is carried out by priority genetic scheduling algorithm based on child. The cost sending data between the tasks equals to L+n/B where n is the delivered data cost, (B) is the bandwidth an (L) is the delay time of the dispatched data. Similarity the scheduling problem can be considered as triple (G, B, L) in which G is the task graph, B is bandwidth and L is the delay time.

In the general state, the algorithm of rewriting graph system can be taken into account as three independent rules: 1: Merging single children, 2: Merging all parents, 3: Replicate Parent Merge.

The rule of merging single children has the lowest cost since it does not involve any condition but a sub-graph pattern (figure 1).

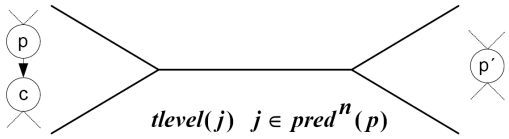


Figure 1. Merging single children

As shown in figure 1, by using this rule we can merge two nodes with each other if in the tasks graph a node has single children and its children owing a single parent. In such conditions, the node C for running will wait the results from node p. If this merging is failed these two nodes may be assigned to two distinct processors in a scheduling and it is obvious that a such action only will lead to delayed entering in to system as large as edge weight connecting between both nodes C and p. By merging two above mentioned nodes the possibility of wrong decisions will be removed when scheduling tasks graph.

At first glance, it seems that if using of an ideal scheduling algorithm taking such implications in to account it can be replaced by above method. However by now such algorithm is not introduced because of high complexity of scheduling. In this paper by distinguishing the problem of tasks graph quality improvement from scheduling one the complexity of the problem has reduced and the possibility of actual implementation has also provided.

The merging of propagated parent includes second merging all parent rules (figure 2). In this rule a task of parent is divided in to a number of tasks based on the number of its child and each of these copies is merged by one of the child. The first row in the list of conditions will be decisive so that the upper bound for each children and their child is not increased by merging copy p.

In the other word if the greatest delay factor of running the nodes c_i is the node p, then by merging it this amount can be reduced (set of nodes C) for each of child having such situation.

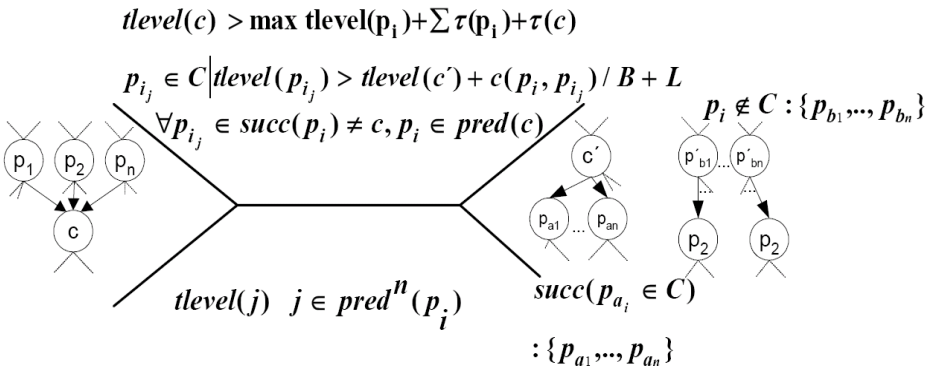


Figure 2. Merging all parents

As shown in figure 2, by using this rule we can If the node p is not the factor affecting the maximum delay of running child c_i, though by merging the node p with nodes c_i the upper bound of that node is not increased, but due to increased volume of running the negative impact of upper bound is transferred to child c_i. Thus, as shown in figure 2, the node p is propagated for the whole child existing within C set and merged in to them; if the node c_i does not exist in C set it will be left alone.

The last rule is the merging of the whole parents as shown in figure 3. Here all of the parents' children of C node denoted by pij are examined. If merging the whole parents with C causes the level of upper bound of these child is not increased the node of interest is added to C set.

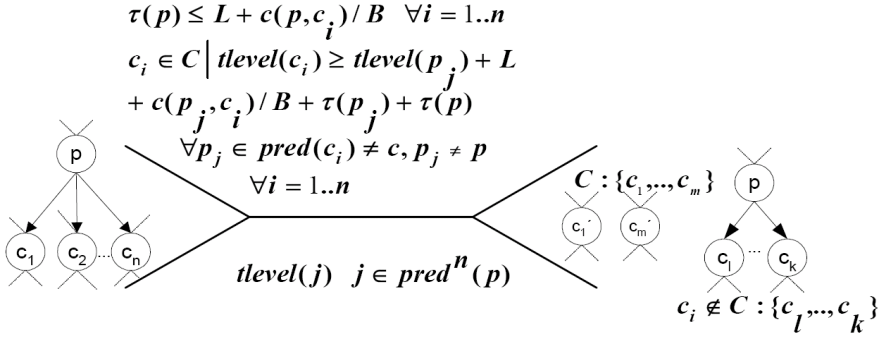


Figure 3. Replicate Parent Merge

The algorithm generating tasks scheduling for initial population in term of the number of offspring is as follows [25]:

a. Arrange the tasks in the order of descending in term of the number of their offspring.
b. Put separately the task with the same number of offspring.
c. Select a task from group of tasks randomly and eliminate it from group.

By using of this algorithm the degree of task graph parallelization will be increased due to decrease the length of critical path. It means that we can obtain a more suitable solution by increasing the number of processors. Having applied the above rules, the graph is prepared for applying the scheduling. The graph resulting from this technique is introduced as an input to prioritized genetic scheduling algorithm based on the number of offspring for scheduling. In the algorithm evaluation section it has shown that by using of this method a suitable scheduling will be established. The proposed method passes the tasks based on prioritizing the new offspring on processors in order to be run. It means that the more the tasks have offspring the sooner scheduling is done. Then that tasks are dedicated to processors P1 to Pm based on EST method in such a way that start time for that task on the processors became short for other processors.

4. SIMULATION AND EVALUATION

In order to evaluate the solution presented in this paper the priority genetic scheduling algorithm based on offspring is used and also the graph resulting from merging is assessed. The scheduling problem is NP-Hard and so far many algorithms have been suggested for it [27,6]. The above optimizing algorithm is applicable before any scheduling, and can be used in place of proposed genetic algorithm.

The solution of the problem or completed time of running the final task in the scheduling algorithm after averaging for any similar number of processors / number of work is shown in table 2. By over viewing the results it is seen that the proposed algorithm has a solution close to optimal one.

By running the proposed algorithms on graph in the figure4, a graph with 16% modification in parallel as well as 6% improvement in length of critical path will be obtained. Also 5%

decrease is seen in number of task graph nodes. Figure 5 shows

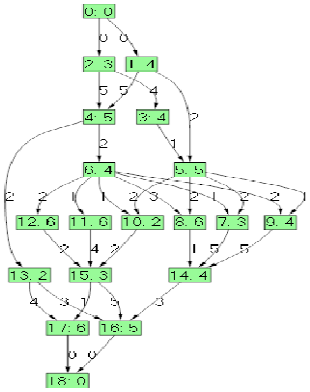


Figure 4. A standard task graph

Table 1. Comparison of the figure 4 graph with existing scheduling algorithm.

MCP	LAST	ETF	HLFET	EZ	LC	NOVIN
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

algorithm stability.

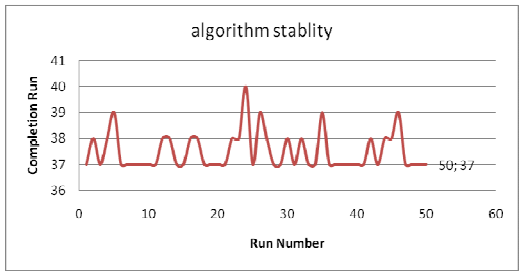


Figure 5. Algorithm stability

This graph is introduced into scheduling algorithm as an input. A run time 38 is obtained with six processors. In the table 1 and figure 6, we see the results of graph figure 4 with algorithms ETF [9], HLFET [7], MCP [8], EZ [16],LC [10].

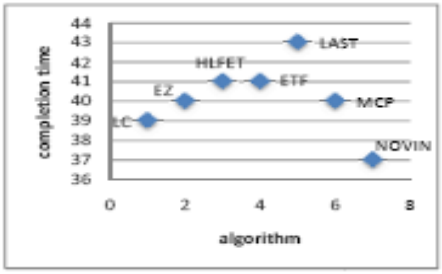


Figure 6. Comparison of the figure 4 graph with existing scheduling algorithm

In order to evaluate the proposed algorithm, a series of simulation has been performed by Visual Basic.net 2005 package on a computer Pentium IV equipped with processor AMD 2.8 MHZ with 512 megabyte memory. By using a formulated program for automatically generating graph, a set of task graphs including 57 graphs with the number of 30, 70, 90 tasks with tasks dependency percentage between 1 to 100 have been established in such a way that the task run time can be varied in the range of 1 to 1000. These graphs were scheduled on multiprocessors with the number 5, 7, 9 ones for genetic algorithm without applying any priority, genetic algorithm based on offspring as well as new proposed algorithms. In order to take other thing being equal, the experiment of all the running steps and operators have been considered the same for all the three algorithms except for prioritizing and establishing initial population. The values of parameters of initial population for algorithms are the number of tasks and replications (number of generations) respectively. Prioritizing based on the number of offspring is better than on the height because it is possible that a task has higher height than the other one and at the same time has more offspring. Logically, running the tasks with more offspring have a greater priority compared to the ones having lower height. It allows more tasks to be run, because its prior task has been done. On the other hand, it is impossible, that the task has more offspring to be the task with less offspring. Therefore, its running does not depend on finishing the task with less offspring.

The solution of the problem or completed time of running the final task in the scheduling algorithm after averaging for any similar number of processors / number of tasks is shown in table 2. By over viewing the results it is seen that the proposed algorithm improves the solution close to optimal one 39% and 24 % compared to priority based on height algorithm and 17 % compared to random genetic algorithm and priority genetic algorithm in term of number of offspring respectively. However, the computing time and implementation of the four algorithms are within the same place since only the parts of the four algorithms differ with each other; each is computed once and initially for generation of primary population.

Table 2. The schedules for last algorithms and proposed algorithm (Merge and Priority based on children)

Algorithms	Total Finish Time (TFT) (seconds)								
	Number of processors = 3			Number of processors = 5			Number of processors = 7		
	Number of tasks			Number of tasks			Number of tasks		
	30	50	70	30	50	70	30	50	70
Algorithm Genetic(A)	1066	2729	7223	2604	2135	3498	855	2314	2919
Priority based on height(B)	978	2440	6003	1882	2005	3222	850	2199	2888
Priority based on children(C)	918	2300	5792	2222	1981	3197	809	2065	2742
Merge and Priority based on children(D)	803	1947	4475	1491	1644	2588	736	1789	2251
Optimal percent (D) to (A)	32	40	61	74	30	35	16	29	30
Optimal percent (D) to (B)	21	25	34	26	21	24	15	22	28
Optimal percent (D) to (C)	14	18	29	5	20	23	10	15	21

Table 2 shows the schedules and TFT for four scheduling algorithms. The results indicate that our suggested new algorithm finds better schedule with minimum TFT compared to the other heuristics. While the computation time of the two above genetic algorithms is more than the other heuristics obviously, and is quite similar, as only their initial population producing step is different, the step is calculated once. As it is shown, while the number of tasks, the number of processors and task implementation within specified time are taken constant.

As the task dependency percentage increases, corresponding to proposed algorithm optimized percentage for obtaining the answer close to optimal one, it increases over the two other algorithms. The main reason of this feather is attributed to merge the tasks prior to prioritize the tasks based on the number of offspring. Since the higher task dependency percentage is taken into account, the more merging of tasks and offspring is and the scheduling became better.

5. CONCLUSION

In this paper, a method is presented for improvement of scheduling algorithm efficiency on graph through technique of merging tasks GRS. First the existing rules in GRS are applied on graph of interest then the output is given to genetic scheduling algorithm based on the number offspring (children). The results from applying the proposed solution on graphs show that after merging the task, the existing tasks within graph become small, the parallel increases and the length of critical path is shorten.

6. ACKNOWLEDGMENT

This work is supported by mahshahr branch islamic azad university and it is a scientific design.

7. REFERENCES

[1] Kirkpatrick, S., Gelatt Jr., C.D., and Vecchi, M.P. 1983. Optimization by simulated annealing, Science 220, 4598 (May 1983), 671_680.

[2] Russell, S. J. and Norvig, P. 1995. Artificial Intelligence: A Modern Approach, Prentice_ Hall, Englewood Cliffs, NJ.

[3] Zomaya, A.Y. and Kazman, R. 1999. Simulated annealing techniques, in "Algorithms and Theory of Computation Handbook" (M. J. Atallah, Ed.), pp. 37-1_37-19, CRC Press, Boca Raton, FL.

[4] Freund, R.F. and Siegel, H.J. 1993. Heterogeneous processing, IEEE Compute. 26.6 (June 1993), 13_17.

[5] Sih, G.C. and E.A. Lee. 1993. A compile- time scheduling heuristic for interconnection- constrained heterogeneous processor architectures, IEEE Trans. Parallel Distrib. Systems 4 (Feb. 1993), 175_186.

[6] Y. K. Kwok and Ahmad, I. 1999. "Static Scheduling Algorithms for Allocating Directed Task Graph Multiprocessors", ACM Cumputing surveys,Vol. December 1999.

[7] Adam, T.L., Chandy, K.M., Dickson, J. 1974. "A Comoarison of List Scheduling for Parallel Processing Systems," communications of the ACM, vol. 17, DEC. 1974, PP. 685-690.

[8] M.-Y. WU, D.D. Gajski, "Hypertool: A Programming aid for Message-Passing Systems," IEEE transactions on parallel and distributed systems, vol. 1, NO. 3, JULY 1990, PP.330-343.

[9] J.J. Hwang, Y.C. Chow, F.D. Anger, C.Y. Lee, "Scheduling Precedence Graphs in Systems with Interproccessor Commonication Times," siam journal on computing, vol. 18, NO. 2, APR. 1989, PP. 244-257.

[10] Kim, S.J. and Brown, J.C. 1988. "A General Approach to Mapping of Prallel Computstion Upon Multiprocessor Architectures," procedings of international conference on parallel processing, vol. II, AUG. 1988, PP. 1-8.

[11] Kruatrachue, B. and Lewis, T. 1988. "Grain Size Determination for Parallel Processing", IEEE 1988.

[12] Freund, R.F., Gherrity, M., Ambrosius, S., Campbell, M., Halderman, M., Hensgen, D., Keith, E., Kidd, T., Kussow, M., Lima, J.D., Mirabile, F., Moore, L., Rust, B., and Siegel, H. J. 1998. Scheduling resources in multi- user,

heterogeneous, computing environments with SmartNet, in :7th IEEE Heterogeneous Computing Workshop (HCW'98), "pp. 184_199, 1998), 13-17.

[13] Shroff, P., Watson, D., Flann, N. and Freund, R. 1996. Genetic simulated annealing for scheduling data- dependent tasks in heterogeneous environments, in "5th IEEE Heterogeneous Computing Workshop (HCW '96)," pp. 98_104.

[14] Armstrong, R., Hensgen, D. and Kidd, T. 1998. The relative performance of various mapping algorithms is independent of sizable variances in run-time predictions, in 7th IEEE Heterogeneous Computing Workshop (HCW, 98), "PP. 79-87.

[15] C.-C. Shen and W.-H. Tsai, A graph matching approach to optimal task assignment in distributed computing system using a minimax criterion, IEEE Trans. Comput. 34, 3 (Mar. 1985), 197_203.

[16] Sarkar, V. 1989. ``Partitioning and Scheduling Parallel Programs for Multiprocessors," MIT press, cambridge, MA.

[17] Singh, H. and Youssef, A. 1996. Mapping and scheduling heterogeneous task graphs using genetic algorithms, in "5th IEEE Heterogeneous Computing Workshop (HCW '96)," pp. 86_97.

[18] Glover, F. and Laguna, M. 1997. Tabu Search, Kluwer Academic, Boston, MA.

[19] Naharai, B., Youssef, A. and Choi, H.A. 1994. Matching and scheduling in a generalized optimal selection theory, in "Proceedings of the Heterogeneous Computing Workshop," pp. 3-8, April 1994.

[20] Lee, Y.H., Chen, C. 2003. A Modified Genetic Algorithm for Task Scheduling in Multi processor Systems, the ninth workshop on compiler techniques for high performance computing.

[21] Shenassa, M.H., Mahmoodi, M. 2006. A novel intelligent method for task scheduling in multiprocessor systems using genetic algorithm, journal of Franklin institute, Elsevier.

[22] Aronosson, P., Fritzson, P. 2002. "Task Merging and Reolization Using Graph Rewriting'.

[23] Ayed, M. and Gaudiot, J.L. 2000. "An Efficint Heuristic for Code Partitioning', parallel computing, 26:399–426.

[24] Shenassa, M.H., Mahmoodi, M. 2005. A new intelligent method for task scheduling in multiprocessor systems using genetic algorithm, Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization, Sharjah, U.A.E. February 1-3, 2005.

[25] Abdeyazdan, M., and Rahmani, A.M. 2008. " Multiprocessor Task Scheduling using a new Prioritizing Genetic Algorithm based on number of Task Children ", 7th International Conference on Distributed and Parallel systems DAPSYS' 2008 Debrecen, Hungary, September 3 – 5, 2008.

[26] Aronosson, P., and Fritzson, P. 2005. "A Task Merging Technique for Paralllization of Modelica Models", March 2005.

[27] Abdeyazdan, M., Rahmani, A.M., Arjman, V., Ghanavati, H.R. 2011. "Genetic Algorithm based on number of children and height task for multiprocessor task Scheduling" WORLD COMP" 11 Springer 2011.

Implementation of Information Retrieval Indonesian Text Documents Using the Vector Space Model

Taqwa Hariguna
Department of informatics
STMIK AMIKOM Purwokerto
Jl.Pol Sumarto Watumas, Purwokerto
085226961748
thariguna@gmail.com

Berlilana
Department of InformationSystem
STMIK AMIKOM Purwokerto
Jl.Pol Sumarto Watumas, Purwokerto
081327222678
berli_amikom@yahoo.co.id

Fandy Setyo Utomo
Department of Information System
STMIK AMIKOM Purwokerto
Jl.Pol Sumarto Watumas, Purwokerto
085281986086
fandy_setyo_utomo@yahoo.com

ABSTRACT
Information search (usually a document) that is based on a query (user input) which is expected to meet user wishes of a collection of documents known as Information Retrieval. This study discusses the implementation of information retrieval to find and match the Indonesian language text documents using the Vector Space Retrieval Model. goal is to provide a solution in search engines to provide information matching text in the database using specific keywords, the result of matching is presented in the form of ratings.

Keyword:
Indonesian Language, Information retrieval, Vector Space Retrieval Model

1. INTRODUCTION
ISO 2382 / 1 defines the Information Retrieval (IR) as acts, methods and procedures to recover the stored data, then provide the required information on the subject. Such actions include text indexing, inquiry analysis, and relevance analysis [2]. The data includes text, tables, images, speech, and video. Including knowledge of relevant information needed to support problem solving and knowledge acquisition. The purpose of the IR system is to meet the needs of users with information to retrieve all documents that may be relevant, at the same time as little as possible to retrieve not relevant documents. This system uses a heuristic function to obtain documents relevant to user queries. A good IR system lets users quickly and accurately determine whether the contents of the documents received meet their needs. In order to better representation of documents, documents with similar topics or content that are grouped together [5].

2. PURPOSE
The purpose of this study is to provide a way to do a search and matching text in a database Indonesian language, previously, the user must be enter a keyword in the search area. The final results of the search text documents will be made in the form of ratings.

3. METHODOLOGY
The methodology used in this study is the Vector Space Retrieval Model. In IR systems, the similarity between documents is defined by bag-of-words representation and converted into a vector space model (vector space model, VSM). This model introduced by Salton [7] and has been used extensively.

In VSM, each document in the database and the user query is represented by a multi-dimensional vector [2, 6]. Dimensions according to the number of terms in the documents involved. In this model:

1. Vocabulary is a collection of all distinct terms remaining after preprocessing of the document and contain the term t index. These terms form a vector space.
2. Each term i in document or query j, given a real-valued weights wij.
3. Documents and queries are expressed as t-dimensional vector dj = (w1, w2, ..., wtj) and there are n documents in the collection, ie j = 1, 2, ..., n.

Examples of three-dimensional vector space models for the two documents D1 and D2, a user query Q1, and three terms T1, T2 and T3 are shown in Figure 2 below,

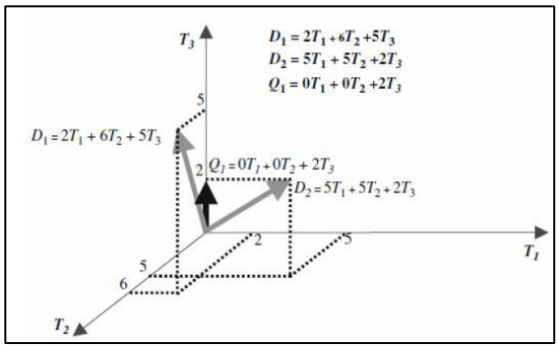


Figure 1: Example of Vector Space Model [2]

In the vector space model, a collection of documents represented by term document matrix (or matrix of term-frequency). Each cell in the matrix corresponds to a given weight of a term in the document specified. A value of zero means that the term is not present in the document. Figure 2 confirm this explanation [2].

$$\begin{bmatrix} & T_1 & T_2 & \dots & T_t \\ D_1 & w_{11} & w_{21} & \dots & w_{t1} \\ D_2 & w_{12} & w_{22} & \dots & w_{t2} \\ \dots & \dots & \dots & \dots & \dots \\ D_n & w_{1n} & w_{2n} & \dots & w_m \end{bmatrix}$$

Figure 2: Example of term-document matrix for the database with n documents and t terms [2]

The success of this model is determined by VSM to a term weighting schemes for both local and global coverage, and the normalization factor [6]. Local weighting is only based on the frequency of appearance of terms in a document and do not see the frequency of appearance of that term in other documents. Approach in the local weighting of the most widely applied is the term frequency (tf) although there are other schemes such as binary-weighted, normalized augmented tf, tf logarithmic and logarithmic alternatives.

Global weighting used to exert pressure on the term that resulted in differences and based on the deployment of certain terms throughout the document. Many schemes based on the consideration that the more often a term appears in the total collection of the term has become more distinct. Use of this weighting can eliminate the need for stop word removal as a stop word has a very small global weights. But in practice better eliminate the stop word in the pre-processing phase, so the fewer terms that must be addressed. Approach to the global weighting include inverse document frequency (idf), squared idf, idf probabilistic, GF-idf, entropy. Idf weighting approach is the most widely used today. Some applications do not involve global weights, just pay attention to tf, when tf is very small or when necessary emphasis on the term frequency within a document [6].

Normalization factor used to normalize the vector so that the document retrieval process is not affected by the length of the document. Normalization is necessary because long documents usually contain the same terms looping thereby increasing term frequency (tf). Lengthy document also contains a lot of different terms that measure the similarity between the query raised by the document, enhance opportunities in retrieving documents that are longer. Several approaches are normalized cosine normalization, the sum of weights, normalized to-4, the maximum weight normalization and pivoted unique normalization.

Local weights of a term i in document j (tfij) can be defined as,

Equation 1. $$tfi_j = \frac{f_{ij}}{\max_i(f_{ij})}$$

Where fij is the number of times term i appears in document j. The frequency is normalized to the frequency of the most common term in the document.

Global weighting of a term i on inverse document frequency approach (idfi) can be defined as,

$$idf_i = \log_2(\frac{n}{df_i})$$

Equation2.

Where DFI is the document frequency of term i and equal to the number of documents containing term i. Log2 is used to minimize the effect relative to tfij.

The weight of term i in the IR system (wij) is computed using the tf-idf measure of the defined as follows [2,3]:

Equation 3. $$w_{ij} = tf_{ij} \times idf_i$$

The highest weighting was given to terms that appear frequently in the document j but rarely in other documents.

Vector space model and tf-idf weighting is used to represent numerical values so that documents can then be calculated proximity between the documents. The closer two vectors in a VSM, the more similar two documents are represented by vectors. The similarity between documents is calculated using a similarity measure function (similarity measure). This size allows the ranking of documents according to similarity (relevance) to the query. After the document is ranked, a fixed number of top-scoring documents are returned to the user. Alternatively, a threshold can be used to decide how many documents will be returned. Threshold can be used to control the push-pull between precision and recall. High threshold value will usually result in high precision and low recall.

One measure of similarity of the popular text is the cosine similarity. This measure calculates the cosine angle between two vectors. If there are two document vectors dj and query q, and t terms extracted from a collection of documents is the cosine between the DJ and q is defined as [2]:

Equation 4. $$similarity\ (\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^{t}(w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}}$$

4. DISCUSSION

In this study, 5 documents in Microsoft Word format (. docx) is used as a source document whose information will be Retrieve. Here is the document name and contents of each document,

Table 1: Document Sources

No	Dokumen	Isi Dokumen
D1	BIN.docx	Komandan Komando Pendidikan dan Latihan TNI Angkatan Darat Letnan Jenderal TNI Marciano Norman ditunjuk oleh Presiden Susilo Bambang Yudhoyono
D2	BUMN.docx	Dahlan Iskan didaulat sebagai Menteri Badan Usaha Milik Negara menggantikan Mustafa Abubakar
D3	gedung dpr.docx	Rencana pembangunan gedung baru DPR yang beberapa waktu lalu menuai kontroversi
D4	Humanoid.docx	Negeri sakura memang pengusung konsep-konsep robot humanoid tercanggih di Asia
D5	Industri.docx	Industri komunikasi dan kolaborasi enterprise di seluruh Asia Pasifik diprediksi berkembang sangat positif pada tahun 2012

4.1. System Design Methodology

In designing the system to process information retrieval there are several methods used are:

4.1.1. Methodology Text Indexing

In text indexing steps are as follows:

a. tokenizing
   Tokenizing is the process of removing punctuation in sentences in the document so as to produce the words that stand respectively.

b. filtering
   Filtering stage is the stage of making important words from the results of tokenizing. This filtering stage uses a list stop list or wordlist. Stop list that is filtering to the words that are not eligible to serve as a differentiator or as a keyword in the search for documents so that these words can be removed from the document. While the wordlist is a list of words that may be used as keywords in the search of documents, so then of course the number of words included in the wordlist will be more than a stop list.

c. stemming
   Stemming is the process of converting a word being said, essentially by removing affixes, affixes to the word in the document. In this study, the process of stemming using Porter's algorithm.
   Here are the steps to the process in the Porter Stemming algorithm [1],
   1) Remove the Particle.
   2) Remove Possessive Pronoun
   3) Remove the first prefix. If no proceed to step 4a, if it exists then go to step 4b.
   4) Delete the second prefix, proceed to step 5a.
      Remove the. If not found then the word assumed as a root word. If found then proceed to step 5b.
   5) Remove the. Then the final word is assumed as a root word. Delete the second prefix. Then the final word is assumed as root word.

Table 2: Rules for Inflectional Particle

Akhiran	Replacement	Additional Condition
-kah	Null	Null
-lah	Null	Null
-pun	Null	Null

Table 3: Rules for Inflectional Possessive Pronoun

Akhiran	Replacement	Additional Condition
-ku	Null	Null
-mu	Null	Null
-nya	Null	Null

Table 4: Rules for First Order Derivational Prefix

Awalan	Replacement	Additional Condition
Meng-	Null	Null
Meny-	S	V…*
Men-	Null	Null
Mem-	P	V…
Mem-	Null	Null

Awalan	Replacement	Additional Condition
Meng-	Null	Null
Meny-	S	V…*
Men-	Null	Null
Mem-	P	V…
Mem-	Null	Null
Me-	Null	Null
Peng-	Null	Null
Peny-	S	V…
Pen-	Null	Null
Pem-	P	V…
Pem-	Null	Null
di-	Null	Null
Ter-	Null	Null
Ke-	Null	Null

Table 5: Rules for Second Order Prefix Derivational

Awalan	Replacement	Additional Condition
Ber-	Null	Null
Bel-	Null	*Ajar*
Be-	Null	*k*er*
Per-	Null	Null
Pel-	Null	*Ajar*
Pe-	Null	Null

Table 6: Rules for Derivational Suffix

Akhiran	Replacement	Additional Condition
-kan	Null	Prefix bukan anggota {ke, peng}
-an	Null	Prefix bukan anggota {di, meng, ter}
-i	Null	Prefix bukan anggota {ber, ke, peng}

d. Indexing
   Text document that has been through the process of tokenizing, filtering, and stemming, then in-indexes into the database.

4.1.2. Searching Methodology (Query)

After indexing text followed by a searching, searching is done as follows:

a. query, Users to search documents by making a search query.

b. stemming, Method for Stemming the same as the processes that exist in text indexing methodology, namely by using the Porter Stemming Algorithm.
c. weighting Document, Calculations for weighting document using Equation 3, ie using TF-IDF algorithm.
d. similarity Calculation, Calculations to measure the level of document similarity (Similarity Calculation) using Equation 4, the Cosine Similarity.
e. Retrieved Document, Documents that have been calculated level of similarity, then presented to the user in the form of the document ranking.

4.2 Analysis

4.2.1. Stage Text Indexing

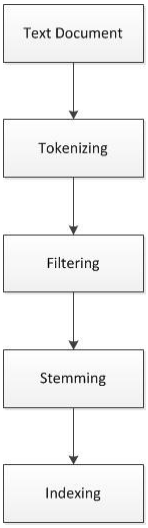


Figure 3 : Indexing Text Methodology

After going through the stages of tokenizing, filtering, and the obtained results stemming indexing every word documents as follows:

Table 7: Results Stemming

No	Dokumen	Hasil Indexing Text
D1	BIN.docx	Komandan komando di latih TN angkat darat letnan jenderal marciano norm tunjuk presiden susilo bambang yudhoyono
D2	BUMN.docx	Dahl Is daulat menteri usaha negara ganti mustafa abubakar
D3	gedung dpr.docx	Rencana bangun DPR ua kontroversi
D4	Humanoid.docx	Neger sakura usung konsep-konsep robot humanoid canggih asia
D5	Industri.docx	Industri komunikasi kolaborasi enterprise asia pasifik prediks kembang positif 2012

4.2.2. Searching Phase (Query)

If there is a query: "Industry communications", then by using TF-IDF algorithm in Equation 3, the calculations can be analyzed to find the weight of each term i in document j, namely as follows:

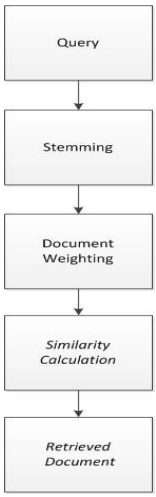


Figure 4 : Searching Methodology

Table 8: Weighting Document

	tf					idf		Wdt=tf.idf				
Term	Q	D1	D2	D3	D4	D5	df	log(n/df)	Q	D1	D2	D3	D4	D5
komandan	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
komando	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
di	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
latih	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
TN	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
angkat	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
darat	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
letnan	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
jenderal	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
marciano	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
norm	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
tunjuk	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
presiden	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
susilo	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
bambang	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
yudhoyono	0	1	0	0	0	0	1	0.69897	0	0.69897	0	0	0	0
Dahl	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
Is	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
daulat	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
menteri	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
usaha	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
negara	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
ganti	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
mustafa	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0

abubakar	0	0	1	0	0	0	1	0.69897	0	0	0.69897	0	0	0
Rencana	0	0	0	1	0	0	1	0.69897	0	0	0	0.69897	0	0
bangun	0	0	0	1	0	0	1	0.69897	0	0	0	0.69897	0	0
DPR	0	0	0	1	0	0	1	0.69897	0	0	0	0.69897	0	0
ua	0	0	0	1	0	0	1	0.69897	0	0	0	0.69897	0	0
kontroversi	0	0	0	1	0	0	1	0.69897	0	0	0	0.69897	0	0
Neger	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
sakura	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
usung	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
konsep-konsep	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
robot	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
humanoid	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
canggih	0	0	0	0	1	0	1	0.69897	0	0	0	0	0.69897	0
asia	0	0	0	0	1	1	2	0.39794	0	0	0	0	0.39794	0.39794
industri	1	0	0	0	0	1	1	0.69897	0.69897	0	0	0	0	0.69897
komunikasi	1	0	0	0	0	1	1	0.69897	0.69897	0	0	0	0	0.69897
kolaborasi	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
enterprise	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
pasifik	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
prediks	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
kembang	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
positif	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
2012	0	0	0	0	0	1	1	0.69897	0	0	0	0	0	0.69897
								TOTAL	1.39794	11.18352	6.29073	3.49485	5.29073	6.68867

From the above data it can be seen the weight of each document is as follows :

Weights for D1: 0 + 0 = 0
Weights for the D2: 0 + 0 = 0
Weights for D3: 0 + 0 = 0
Weights for the D4: 0 + 0 = 0
Weights for D5: 0.69897 + 0.69897 = 1.39794.

After the weights of each document is found, then look for Cosine Similarity values using equation 4, to determine the degree of similarity of documents that exist in the database with the specified query,

Table 9: Calculation of Vector Space Model

Term	Query	D1	D2	D3	D4	D5	Q*D1	Q*D2	Q*D3	Q*D4	Q*D5
komandan	0	0.48856	0	0	0	0	0	0	0	0	0
komando	0	0.48856	0	0	0	0	0	0	0	0	0
di	0	0.48856	0	0	0	0	0	0	0	0	0
latih	0	0.48856	0	0	0	0	0	0	0	0	0
TN	0	0.48856	0	0	0	0	0	0	0	0	0
angkat	0	0.48856	0	0	0	0	0	0	0	0	0
darat	0	0.48856	0	0	0	0	0	0	0	0	0
lenan	0	0.48856	0	0	0	0	0	0	0	0	0
jenderal	0	0.48856	0	0	0	0	0	0	0	0	0
marciano	0	0.48856	0	0	0	0	0	0	0	0	0
norm	0	0.48856	0	0	0	0	0	0	0	0	0
tunjuk	0	0.48856	0	0	0	0	0	0	0	0	0
presiden	0	0.48856	0	0	0	0	0	0	0	0	0
susilo	0	0.48856	0	0	0	0	0	0	0	0	0
bambang	0	0.48856	0	0	0	0	0	0	0	0	0
yudhoyono	0	0.48856	0	0	0	0	0	0	0	0	0
Dubl	0	0	0.48856	0	0	0	0	0	0	0	0
Is	0	0	0.48856	0	0	0	0	0	0	0	0
dualat	0	0	0.48856	0	0	0	0	0	0	0	0
menteri	0	0	0.48856	0	0	0	0	0	0	0	0
usaha	0	0	0.48856	0	0	0	0	0	0	0	0
negara	0	0	0.48856	0	0	0	0	0	0	0	0
ganti	0	0	0.48856	0	0	0	0	0	0	0	0
mustafa	0	0	0.48856	0	0	0	0	0	0	0	0
abubakar	0	0	0.48856	0	0	0	0	0	0	0	0
Rencana	0	0	0	0.48856	0	0	0	0	0	0	0
bangun	0	0	0	0.48856	0	0	0	0	0	0	0
DPR	0	0	0	0.48856	0	0	0	0	0	0	0
ua	0	0	0	0.48856	0	0	0	0	0	0	0
kontroversi	0	0	0	0.48856	0	0	0	0	0	0	0
Neger	0	0	0	0	0.48856	0	0	0	0	0	0
usung	0	0	0	0	0.48856	0	0	0	0	0	0
konsep-konsep	0	0	0	0	0.48856	0	0	0	0	0	0
robot	0	0	0	0	0.48856	0	0	0	0	0	0
humanoid	0	0	0	0	0.48856	0	0	0	0	0	0
canggih	0	0	0	0	0.48856	0	0	0	0	0	0
asia	0	0	0	0	0.15836	0.15836	0	0	0	0	0
industri	0.48856	0	0	0	0	0.48856	0	0	0	0	0.23869
komunikasi	0.48856	0	0	0	0	0.48856	0	0	0	0	0.23869
kolaborasi	0	0	0	0	0	0.48856	0	0	0	0	0
enterprise	0	0	0	0	0	0.48856	0	0	0	0	0
pasifik	0	0	0	0	0	0.48856	0	0	0	0	0
prediks	0	0	0	0	0	0.48856	0	0	0	0	0
kembang	0	0	0	0	0	0.48856	0	0	0	0	0
positif	0	0	0	0	0	0.48856	0	0	0	0	0
2012	0	0	0	0	0	0.48856	0	0	0	0	0
	Sqrt(Q)			Sqrt(D_t)						Sqrt($Q*D_t$)	
	0.98849	2.79588	2.09691	1.56295	1.89163	2.13434	0	0	0	0	0.69093

To calculate the cosine angle between the vector query with each document can use the formula:

Cosine (In) = Sqrt (Q * At) / (Sqrt (Q) * Sqrt (In))
Document 1 (D1)
Cosine (D1) = 0 / (0.98849 * 2.79588) = 0
Document 2 (D2)
Cosine (D2) = 0 / (0.98849 * 2.09691) = 0
Document 3 (D3)
Cosine (D3) = 0 / (0.98849 * 1.56295) = 0
Document 4 (D4)
Cosine (D4) = 0 / (0.98849 * 1.89163) = 0
Document 5 (D5)
Cosine (D5) = 0.69093 / (0.98849 * 2.13434) = 0.32749.

From the calculation above, can be ranked as the level of similarity with the query document "Industry Communications", is as follows : D5, D1, D2, D3, D4. Can be seen that, the document 5 (D5) has a high level of similarity with the query.

5. CONCLUSION

Information Retrieval in Document Indonesian Text Retrieval Using the Vector Space Model can provide a solution in search engines to provide information matching text in the database using specific keywords, the result of matching is presented in the form of ratings. This study used only for Indonesian language vocabulary, if will be used for multi-lingual English and Indonesia at the Steaming can be removed or replaced with other stemming algorithms such as porter stemming algorithm for English.

6. REFERENCES

[1]. Agusta, L. 2009. Perbandingan Algoritma Stemming Porter dengan Algoritma Nazief dan Adriani untuk Stemming Dokumen Teks Bahasa Indonesia. Pada Konferensi Nasional Sistem dan Informatika 2009. No Jurnal : KNS&I09-036.

[2]. Cios, K.J. Etc. 2007. Data Mining A Knowledge Discovery Approach, Springer.

[3]. Lee, D.L. 1997. Document Ranking and the Vector-Space Model. IEEE March-April 1997.

[4]. Lu, G. 1999. Multimedia Database Management Systems. ARTECH HOUSE Inc. Canton Street.

[5]. Murad, Azmi MA., Martin, Trevor. 2007. Word Similarity for Document Gouping using Soft Computing. IJCSNS International Journal of Computer

Science and Network Security, Vol.7 No.8, August 2007, pp. 20- 27

[6]. Polettini, N. 2004. The Vector Space Model in Information Retrieval – Term Weighting Problem

[7]. Salton, G. 1983. Introduction to Modern Information Retrieval, McGraw Hill.

Compression and Decompression Application for HTML Script Files

Nulita
Informatics Department,
Faculty of Engineering and Computer Science
Krida Wacana Christian University
Tanjung Duren Raya No. 4, Jakarta 11470
021-5666952

Nina Sevani
Informatics Department
Faculty of Engineering and Computer Science
Krida Wacana Christian University
Tanjung Duren Raya No.4, Jakarta 11470
021-5666952
nina.sevani@ukrida.ac.id

ABSTRACT
The complexity of function and interface in a website increases the size of HTML's files. As a result, it needs bigger bandwidth and more time for transmitting the file between client and server. Using Lempel Ziv Welch (LZW) algorithm, the size of the HTML file can be reduced without making any changes in the original interface. LZW algorithm works by looking for and comparing each character in the file and changes it to ASCII code based on particular calculation. In the receiver side, the compressed file needs to be decompressed to return to the original size of the compressed file. In the decompression process ASCII code is changed into the original character based on the same calculation. The result of the experiment shows that the compressed file has smaller size and can be transmitted faster than the original file.

Keywords
HTML file, Lempel Ziv Welch, compression, decompression, ASCII code

1. INTRODUCTION
Internet is widely used to spread and transfer data and information around the world. Website is one of the internet services that is used by many people for several purposes, like browsing, trading, recreation, education, and so on. As a result, website becomes more complex in function and interface, featuring multimedia, such as picture, video, sound, animation, and song.

The multimedia makes the size of the file bigger and influences the speed of transfer and access time for the website. This condition is getting worse when the internet backbone is crowded, for example at the office hour. This delayed connection can degrade the website performance.

One of the solutions to overcome the size of the website is by compressing the website file, written in HTML, before it is transmitted to the internet. The compression should not change the original interface of the website. Among compression algorithms, such as Huffman algorithm and Run Length Encoding (RLE) algorithm, Lempel Ziv Welch (LZW) algorithm is more suitable for compressing HTML files because it produces the ratio of the compression file and the compression time better than the others do. It is also the most suitable for general use [1].

This research intends to explain the compression and decompression HTML file, to reduce the size before it is transmitted to the internet, without changing the original interface of the website. This experiment result, show that using the LZW algorithm for compress and decompress HTML file before transmitted can reduce the file size. In the end its can increased transfer speed and shorter time to access the website [2].

2. LITERATURE REVIEW
Some of theories used in this research will be explained briefly in this section, including the theory of compression and decompression file, as well as how LZW algorithm works.

2.1 Compression and Decompression
Compression is a process of changing a group of data into particular codes for saving the space and accelerate the time of data transmission. Figure 1 illustrates the working process of compression.

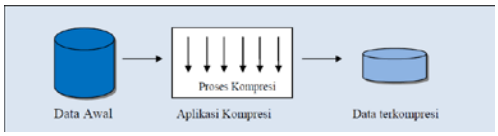


Figure 1. Illustration of Compression Process

There are two processes in the data compression, that are encoding and decoding. Encoding is a process to code data into a code, based on some formats. Decoding is the opposite process of encoding. In decoding, some codes are changed into the original data using the compression algorithm.

Decompression is a process that works in the opposite way from the compression process. In decompression, the compressed file is changed into the original data, as shown in Figure 2.

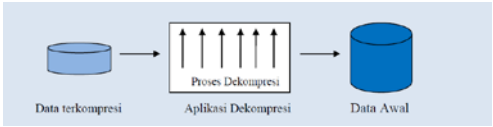


Figure 2. Illustration of Decompression Process

There are two categories of compression: lossless compression and loosely compression. The lossless compression results in the higher compression degree and it can return the compressed file into the original file without any changes. This category fits for digital file and picture. The loosely compression has the smaller compression degree and the compressed file cannot return into the original file. This category fits for database, documents, and spreadsheet [3].

2.2 LZW Algorithm

LZW algorithm is one of compression algorithms that use the dynamic method, making the code map change from time to time. LZW algorithm can adapt with file characteristics upon the compression process [4]. It means that the total amount of a character coded depends on the character that previously appears in the file.

LZW algorithm works using dictionary, where the fragment texts are changed into indexes from the dictionary. Using this algorithm, we can assure that there is no missing data upon the compression process. Here are the steps of encoding from the compression process using LZW algorithm.
Step 1 : Dictionary initialization using all base character {'A', 'B', …, 'Z', 'a', 'b', …, '0', .., '9'}.
Step 2 : Define P that save first character in character stream.

Step 3 : Define C that save next character in character stream.

Step 4 : Check if the string (P+C) is in dictionary

If the answer is Yes, change the value of P became P = P+C.

If the answer is No, replace P with an output code and add string (P+C) in dictionary, give unique number or code for that string. And now replace P with C (P=C).

Step 5 : Check if the character stream still contains character

If the answer is Yes, than go to step 2. If the answer is No, replace string P with output code.

The decoding process works in the opposite way. Start with the dictionary initialization with all base characters, the decompression process works in six steps [5].
Step 1 : Dictionary initialization using all base character {'A', 'B', …, 'Z', 'a', 'b', …, '0', .., '9'}.
Step 2 : Define CW (correct code word) as the first code word in stream code. Look at dictionary and output string CW to charstream.

Step 3 : Define PW (previous code word) as CW.

Step 4 : Define CW as the next code word into stream code.

Step 5 : Check if string CW is in the dictionary?

If the answer is Yes, then output string CW goes to the stream code, define P as string CW and define C as the first character from string CW, and add string (P+C) to dictionary.

If the answer is No, than define P as string PW and C as first character from string PW, and output string (P+C) to stream character and save it to the dictionary. The dictionary is now same with CW.

Step 6 : Check if the stream code contains many code words. If the answer is Yes, go to step 3.

3. APPLICATION DESIGN

This chapter describes the design of the application. There are some software programs used to create this application, namely Microsoft Visual Basic 6.0, Macromedia Dreamweaver MX 2004, Macromedia Firework MX 2004, HTML Help Workshop, and Package and Development Wizard.

3.1 Compression Process

Basically, there are 2 main processes in this application, that are compression and decompression of HTML file. Figure 2 below describes the flowchart of the compression process.

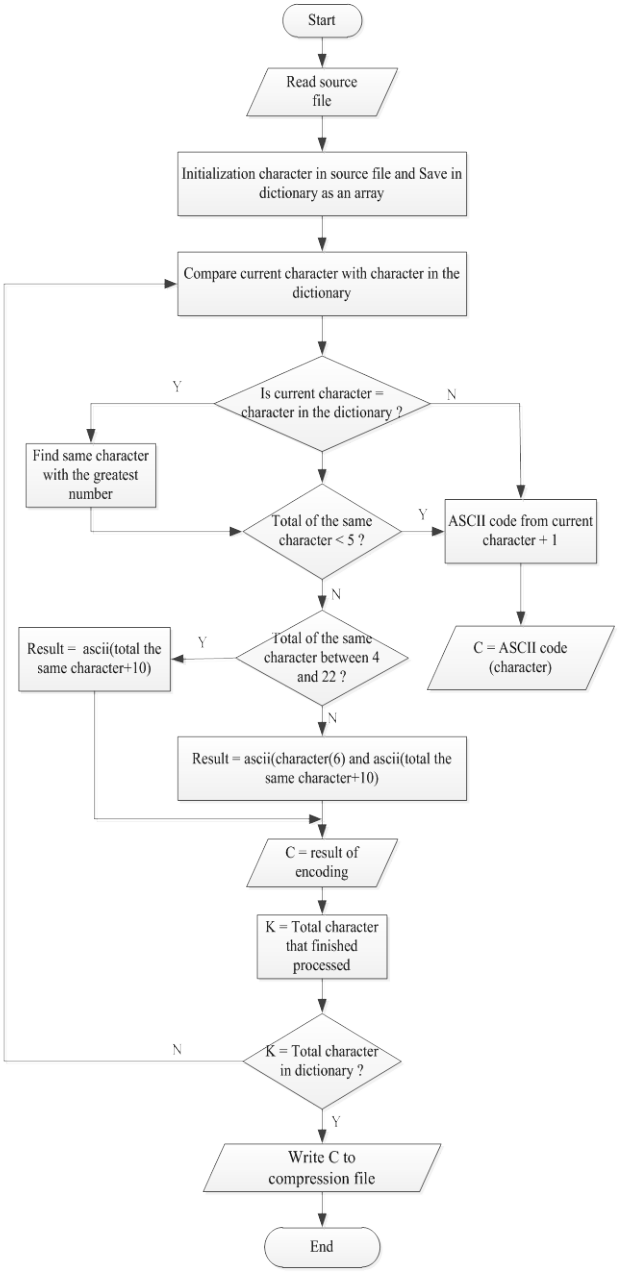


Figure 3. Flowchart of Compression Process

As an example, we will use this part of HTML coding through the compression process using the flowchart in figure 3.

< T I T L E > T E S T I N G … < \ T I T L E >

0 1 2 3 4 5 6 … …. … … …. 94,95,96 (posisi)

Step 1 : Compare characters, start with character at first position.

According to the example above, the character "T" at 3rd position is same as the character at 1st position. After that, compare the 2nd position character, that is "I" with the 4th position character "L". Because the character "I" is not same as the character "L", the comparison process is continued to the next character. Repeat this step until all of characters are processed.

Step 2 : Count the amount of the same character.

In the step 1, we know that the character "T" at 94th position is the same character at 1st, 3rd, 7th, 10th , 96th position. Because the amount of the same character is between 4 and 22, the 94th until 99th position will be encoded as 3 characters. For example:

character 1 = ASCII code(6+10) = ASCII code(16)

character 2 = ASCII code(mod(((94 – 1)/90)+35) = ASCII code(38)

Here are the results of compression process from the part of HTML coding described before :

= U J UMF ? U F T U J O H … ►&$

0 1 2 3 4 5 6 … …. … … …. 94,95,96 (posisi)

3.2 Decompression Process

On the other side, users need to decompress HTML files to see the original interface of the website. Figure 4 describes the flowchart of the decompression process.

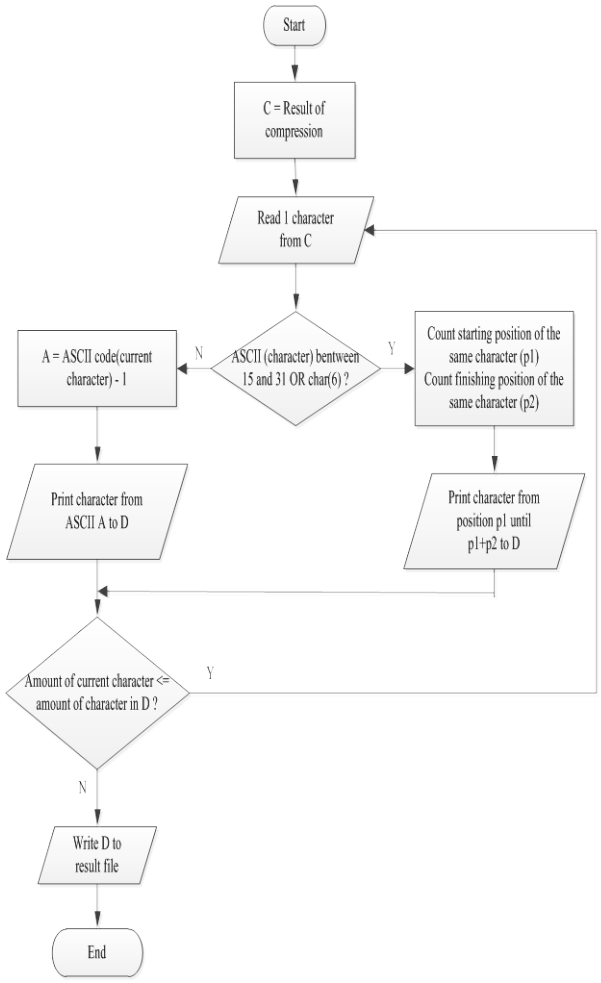


Figure 4. Flowchart of Decompression Process

The decompression process works in the opposite way of the compression process. Using the result of the compression process above, we will observe the steps of the decompression process. Let us assume that D is a compressed file, consisting of the ASCII codes of the original character.

Step 1: Compare each character in D to find if the ASCII code is 6, or between 14 and 32.

If the ASCII code is between 14 until 32, it shows that the character appears 5 until 21 times in the file. We can count the number of the same characters in the file by subtracting the ASCII code by 10.

If the ASCII code is not between 14 and 32, to find the ASCII code of the original character, we can subtract ASCII code by 1.

If the ASCII code is 6, we know that there are more than the same 21 characters in the file, starting from that character.

Step 2: Repeat the step 1 until all characters in the file are processed.

For example, we know that ASCII for character '=' is 61. Because the ASCII code is not between 14 and 32, to know the ASCII code for the original character, we need to subtract the ASCII

code by 1, and get ASCII code 60. The character for ASCII code 60 is '<'. Meanwhile, we know that the ASCII code for character '►' is 16, that is between 14 and 32. We count the ASCII code for the original character by subtracting 16 with 10 and get 6 as the result. This means there are the same 6 characters in the file.

4. ANALYSIS AND RESULT

4.1 Application Interface

The interface of this application is designed to be as user-friendly as possible. The interface consists of button, label, texts, and picture as the symbol of the application. Button is used for adding the file to be processed, deleting and editing file, saving file, and displaying the original file. Figure 5 below shows one of application interfaces. In this figure, we can see a message box to inform users that the compression process is successful. This application also informs the users about: source file size, compressed file size, times needed to do the compression/decompression process, and ratio in percentage about the original and result file.

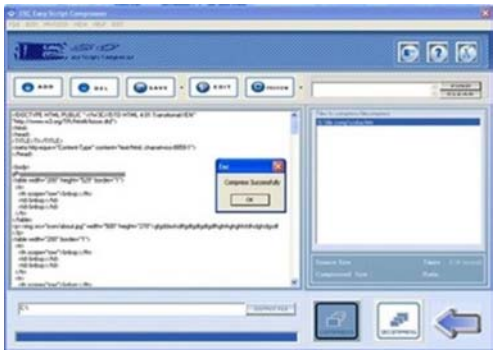


Figure 5. Compression Process Interface

4.2 Comparison of File Size

Table 1 displays the experiment result of file size testing before and after the compression process.

Table 1. Experiment Result of Compressed File Size

Nama *File*	Ukuran *File* (kb)		Rasio (%)	Waktu (Sec)
	Asli	Setelah kompresi		
acehoomegazen_blogspot_com.htm	246.68	76.00	69.13	60.28
TT *Software* « HIDAYAT Blog's.htm	134.54	50.00	62.84	38.95
Search.html	99.99	23.00	77.40	16.86
entity-relationship-diagram-erd.html	79.88	32.00	59.57	24.83
News_Detail.php	53.04	13.00	75.22	10.11

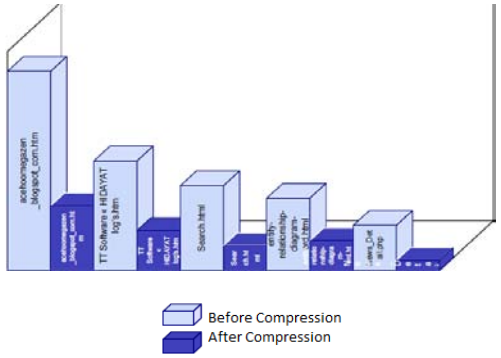


Figure 6. displays the comparison between the file size before and after the compression process.

From Table 1 and Figure 6, we can see that the compression process can reduce the file size. The time needed for the compression process is dependent on the original size of the file.

Compression ratio is a percentage reduction between file size before and after the compression process. Compression ratio is counted by the formula:

```
Ratio(%)= 100-(100*(compressed file size/original
file size))          (1)
```

Table 2 shows the experiment result of the file size before and after the decompression result.

Table 2. Experiment Result of Decompression File Size

Nama *File*	Ukuran *File* (kb)		Rasio (%)	Waktu (Sec)
	File kompresi	*File* dekompresi		
acehoomegazen_blogspot_com.htm	76.00	246	69.10	17.53
TT *Software* « HIDAYAT Blog's.htm	50	130	61.93	5.41
Search.html	22.60	77.00	71.46	2.34
entity-relationship-diagram-erd.html	32.30	78.00	59.25	2.88
News_Detail.php	13.14	48.00	73.83	1.33

Figure 7 displays the comparison between the file size before and after the decompression process.

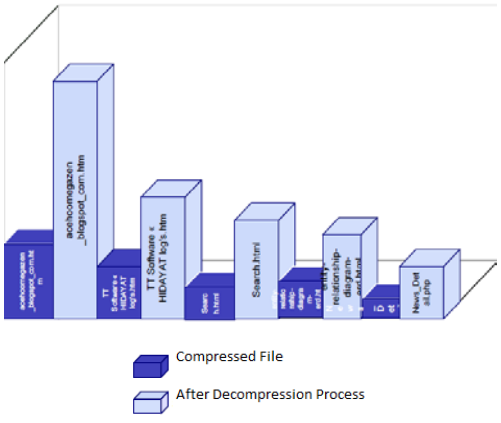


Figure 7. The Comparison Graph of File Size

Based on the data in Table 2 and the graph in Figure 7, we can see that after the decompression process, the file will return to the original file size. The decompression ratio is counted in percentage using this formula:

```
Ratio(%)=100-(100*(decompressed file size/original file size))        (2)
```

5. CONCLUSION

This experiment result shows that the LZW algorithm can be used for compressing some scripts, like htm, html, and php. In this respect, more similar characters with the same length will need longer time to compress and decompress, but they will enhance the ratio.

We can also see that the time needed for the compression process is the same as that needed for the decompression process. The longer for compression the longer for decompression will be. The time needed for the compression and decompression process dynamically changes, depending on the working time of the CPU.

6. REFERENCES

[1] Merdian, M. and Indarto, W. 2005. Implementation of Run Length, Half Byte, and Huffman Algorithm for File Compression. In Proceedings of the SNATI National Conference of Information Technology Application (Yogyakarta, Indonesia, June 18, 2005). H79-H84. DOI = http://journal.uii.ac.id/index.php/Snati/article/view/1391

[2] Kandaga, T. 2006. Implementation Ananlysis of Compression and Decompression Data Using Statistic Method and Dictionary. Journal of Informatics.Volume 2, Number 2. (Bandung, Indonesia, December 2006). 81-91. DOI= http://majour.maranatha.edu/index.php/jurnal-informatika/article/view/253

[3] Mengyi Pu, Ida. 2006. Fundamental Data Compression Book (London, UK). Elsevier.

[4] Linawaty and Panggabean, H. 2004. Performance Comparation Between Huffman, LZW, and DMC Compression Algorithm for Various File Type. Integral. Volume 9 Number 1.

[5] Razi, F. 2009. Influence Analysis of Bits Length in Compression Program Performance Using Lempel Ziv Welch. Undergraduate Thesis. University of North Sumatera.

E-Commerce Technology Adoption by Small Medium Enterprises (SMEs) Case Study: SMEs in Jabodetabek, Indonesia

Nunung Nurul Qomariyah
Program Studi Teknik Informatika Universitas Pembangunan Jaya
Jl. Boulevard Bintaro Sektor 7 - Tangerang Selatan – Indonesia
(62)21-745- 5555
nunung.nq@upj.ac.id

ABSTRACT
The use of e-commerce in small medium-sized enterprises (SMEs) has become an important topic in information system researches. This study focused on SMEs because they play a strategic role to national economic development in Indonesia. The development of the SMEs can gain maximum advantages if the information technology and the e-commerce technology can be utilized.

The main question to be answered by this research is: Do the characteristics of the SMEs in Indonesia that have adopted e-commerce differ from those the non-adopters?

Utilizing a portion of Rogers model of innovation diffusion as the framework and treating e-commerce as a form of new innovation, this study analyze factors affecting e-commerce diffusion combined with some other factor which often discussed in latest researches of Indonesian SMEs. Quantitative research used as a method in this research, by distributing a questionnaire survey to the respondents. A Mann-Whitney U-Test analysis, which is commonly used in nonparametric statistics for two independent samples, was carried out to found significant factors that differentiated adopters and non-adopters of e-commerce.

As a result, I found that one factor of the six, organizational resources and readiness, appear very highly significant affecting adoption of e-commerce by SMEs. Otherwise the five other factors show the good significance level.

Keywords
Electronics commerce adoption, model of innovation adoption, Small Medium Enterprises (SMEs)

1. INTRODUCTION
Development of e-commerce technology today is rapidly along with the development of internet technology in the world. Ease of delivery of information and communication through the internet is then pushed businesses to consider various business opportunities and benefits that may be obtained by the adoption of e-commerce technology. Why this study focused on Small Medium-sized Enterprises (SMEs)? That was because SMEs in Indonesia had a strategic role for national economic development. It is not only increasing economic growth and employment absorptions, but also growing distribution of development outcomes. In the economic crisis, that has occurred over the past decade, many large-scale businesses has remained stagnant and even worse, they discontinued their business, but the SME sector has proved to be more resilient in dealing with crisis situation. The use of information technology and application of e-commerce can be utilized for enforcing SMEs capability [1].

2. RESEARCH LIMITATION
This study examined e-commerce adoption by SMEs. The term e-commerce is more broadly interpreted as a transaction exchange of goods or services rather than exchange of information electronically. According to Chaffey's definition of e-commerce, a static website can also be categorized as e-commerce, because it contains information that can be delivered from the organization to some of its external stakeholders. In this study, the definition of e-commerce refers to what have been described by Chaffey [2]. I used this definition for selecting the proper respondents. The sample size of this study is only 20 of about 200 population size, therefore it can not be claimed that the results were generalizable to all SMEs in Indonesia. This study was limited to SMEs located in Jakarta, Bogor, Depok and Tangerang (JABODETABEK) Indonesia. The data of respondents were obtained from SMESCO (Small and Medium Enterprises and Cooperatives), Indonesia Ministry of Cooperatives and SMEs.

3. LITERATUR REVIEW
3.1 Definition of SMEs
There are many different definitions of SMEs in each country depending on their existing economic and political conditions. Indonesia has the legal definition of SMEs that was clearly mentioned in the Republic of Indonesia Law No.20 Year 2008. The small-sized businesses is the business entity that has the following criterias: (1) has a net worth of more than Rp 50,000,000.00 (fifty million rupiahs) up to a maximum of Rp 500,000,000.00 (five hundred million rupiah) does not include land and building, and (2) has annual sales of more than Rp 300,000,000.00 (three hundred million rupiahs) up to a maximum of Rp 2,500,000,000.00 (two billion five hundred million rupiahs). Meanwhile, the medium-sized businesses is the business entity that has the following criterias: (1) has a net worth of more than Rp 500,000,000.00 (five hundred million rupiahs) up to a maximum of Rp 10,000,000,000.00 (ten billion rupiahs) excluding land and building, and (2) has annual sales of more than Rp 2,500,000,000.00 (two billion five hundred million rupiahs) up to a maximum of Rp 50,000,000,000.00 (fifty billion rupiahs). This definition had become reference to determine the sampling data of this study.

3.2 Characteristics of SMEs in Indonesia
There was a study conducted by Oswari et al [3] that mentioned the major weaknesses of SMEs in Indonesia, they were market

orientation, quality of human resources, capability to use technology, market access and capital. Detailed of the five main weaknesses were described as follows:

1. Most of the products of SMEs did not have the penetration power into the world market (export).
2. Human resources involved in the business of SMEs had been neither qualified nor professional, especially in the field of entrepreneurship. Most of SMEs did not meet market demands due to unstability of quality, hugely number of orders, slowly and lately delivery of goods or services.
3. SMEs had limited capability to use the production process technology, product innovation and product quality knowledge. Most of them still used production technology that was quite simply and traditionally.
4. SMEs had limited abilities and aggressiveness to reach market access.
5. SMEs capital capabilities and abilites to access source of funding was limited due to the lack of collateral, high cost of consulting fees, and high cost of sales promotion (i.e. exhibitions, brochures, the cost of quality testing, shipment of samples, etc). This fact was worsened by inconsistent support from banking and financial sectors.

There were several researches found the level of IT adoption among SMEs in Indonesia was still low, the biggest factors came from internal organizations such as financial and human resources [4].

3.3 Rogers Innovation Adoption Theory

Adoption of innovations theory formulated by Rogers is widely cited in researches involving the renewal of the old system to a new process or new technique. This theory was first issued in 1962, written in a book titled "Diffusion of Innovations" and then continued to be improved in the 2005 edition.

Diffusion is a process by which an innovation is communicated via channel in the spesific period to the members of the social system [5]. Diffusion could be interpreted as a form of communication that contains a new idea as the message. Rogers said there were four main elements in the diffusion of new ideas as follows:

1. Innovation

An innovation is an idea, practice, or object that is perceived as new by an individual or other unit of adoption. The characteristics of an innovation, as perceived by the members of a social system, determine its rate of adoption. An innovation has certain characteristics that affect the speed of adoption. Some of these characteristics are relative advantage, compatibility, complexity, trialability, observability.

2. Communication channels

Communication is the process by which participants create and share information with one another in order to reach a mutual understanding. A communication channel is the means by which messages get from one individual to another. Mass media channels are more effective in creating knowledge of innovations, whereas interpersonal channels are more effective in forming and changing attitudes toward a new idea, and thus in influencing the decision to adopt or reject a new idea. Most individuals evaluate an innovation, not on the basis of scientific research by experts, but through the subjective evaluations of near-peers who have adopted the innovation.

3. Time

The time dimension is involved in diffusion in three ways. First, time is involved in the innovation-decision process. The second way in which time is involved in diffusion is in the innovativeness of an individual or other unit of adoption. The third way in which time is involved in diffusion is in rate of adoption

4. Social system

A social system is defined as a set of interrelated units that are engaged in joint problem-solving to accomplish a common goal.

Rogers wrote that there were five characteristics determined the speed of adoption of an innovation as follows:

1. Relative advantage

Relative advantage is the degree to which an innovation is perceived as better than the idea it supersedes. The degree of relative advantage may be measured in economic terms, but social prestige, convenience, and satisfaction are also important factors. It does not matter so much if an innovation has a great deal of objective advantage. What does matter is whether an individual perceives the innovation as advantageous. The greater the perceived relative advantage of an innovation, the more rapid its rate of adoption will be.

2. Compatibility

Compatibility is the degree to which an innovation is perceived as being consistent with the existing values, past experiences, and needs of potential adopters. An idea that is incompatible with the values and norms of a social system will not be adopted as rapidly as an innovation that is compatible. The adoption of an incompatible innovation often requires the prior adoption of a new value system, which is a relatively slow process.

3. Complexity

Complexity is the degree to which an innovation is perceived as difficult to understand and use. Some innovations are readily understood by most members of a social system; others are more complicated and will be adopted more slowly. New ideas that are simpler to understand are adopted more rapidly than innovations that require the adopter to develop new skills and understandings.

4. Trialability

Trialability is the degree to which an innovation may be experimented with on a limited basis. New ideas that can be tried on the installment plan will generally be adopted more quickly than innovations that are not divisible. An innovation that is trialable represents less uncertainty to the individual who is considering it for adoption, who can learn by doing.

5. Observability

Observability is the degree to which the results of an innovation are visible to others. The easier it is for individuals to see the results of an innovation, the more likely they are to adopt it. Such visibility stimulates peer discussion of a new idea, as friends and neighbors of an adopter often request innovation-evaluation information about it.

4. PREVIOUS STUDIES

There were several previous studies related to this research topic. Some related researches were conducted in Indonesia, Malaysia, Singapore, Canada and Thailand, from the year 2001 to the most recent research in 2009. Those previous researches was cited as reference because the model used were likely similar.

The most recent research was conducted by Kurnia et al [6] which proposed a theoretical model of e-commerce technologies adoption by SMEs in developing countries by incorporating various readiness factors. It then explored the influence of these factors on the adoption of different e-Commerce technologies such as email, internet, intranet, extranet, EDI, EFT and barcode. The environmental pressure is also considered in that study.

Research on the same topic was conducted in Central Okanagon, Canada by Sparling, Toleman and Cater-Steel [7] which was used a survey of SMEs to determined what reasons influenced the low level of e-commerce adoption. Constructs used in the survey focussed on three contexts: organizational, external environmental and innovation. The study found significant factors that differentiated adopters and non-adopters of e-commerce included technological opportunism and readiness, owner experience with computers, support within the organization, relative advantage and compatibility.

Other similar studies were conducted in New Zealand [8] Singapore [9] and Thailand [10] that used the five factors of Rogers' innovation theory as the model of SMEs' e-commerce adoption. And then they determined what factors was influenced the adoption process.

Previous researches in Indonesia were conducted in Pekanbaru [11], Yogyakarta [4] and Depok [12] which was aimed to identify the inhibiting and driving factors for SMEs' e-commerce adoptions in their respective regions.

5. METHODOLOGY

5.1 Model Development

In this research, I developed a model of e-commerce adoption that comprised the three most potential determinants of e-commerce adoption in SMEs from the eight previous researches that used as reference, they are: (1) the resources and readiness of the organization, (2) management support and (3) environmental pressure. These three factors combined with the three other factors from the characteristics of innovation Rogers. In the context of innovation characteristics there are three factors: (1) relative advantage, (2) complexity and (3) trialability.

The model that was used to test SMEs' e-commerce adoption in this study was shown in Figure 1:

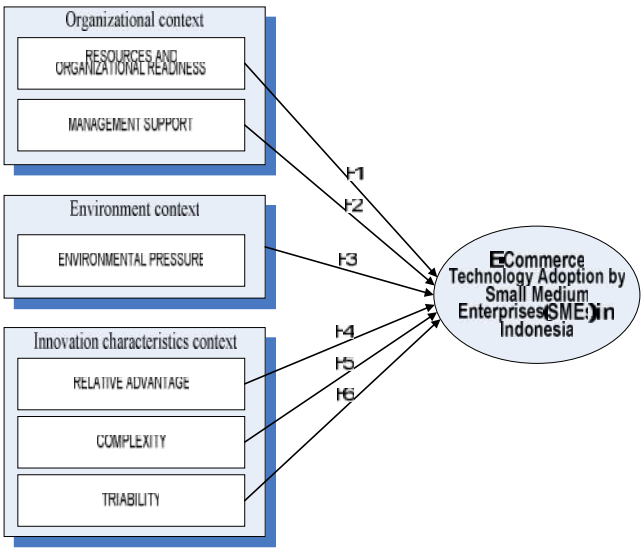


Figure 1. E-commerce Adoption Model of SMEs in Indonesia

After formulating the model, I prepared a questionnaire based on it. Then the questionnaire was tested to the respondent to found out were there any differences in the characteristics of SMEs that have adopted e-commerce as compared to the non-adopters? By highlighting these differences, the result of this study can be used to help SMEs for increasing their e-business initiatives.

Each of these factors was described as follows:

1. Resources and Organizational Readiness

This term referred to the top level manager perceptions of e-commerce technologies and their perceptions of the availability of resources, rules and procedures within the organization that supported adoption of these technologies. Some SMEs especially in developing countries tend to have a low level of organizational readiness due to their lack of resources, infrastructure and procedures that are necessary for e-commerce technologies adoption [6].

2. Management Support

Some related researches showed that senior management support provided a crucial role in the adoption of new technological innovations [7] [10] [8].

3. Environmental Pressure

Environmental context consist of external and internal pressure. Typically the external environmental pressure includes market conditions such as competitive market forces, market uncertainty, and government regulation. Then the internal environmental pressure came from institutional. There are two components of institutional pressure: stakeholder pressure and competitive pressure. Stakeholder pressure came from customers, trading partners, investors, suppliers, media and employees. Competitive pressure forces a business to adopt a technology to gain competitive advantage. As more competitors and trading partners adopt an innovation, small firms are more inclined to adopt the innovation in order to maintain their own competitive position.

4. Relative Advantage

The indicators used for examining these factors are: lower business costs, speed up the completion of work, reach a broader market, preference to upgrade other business ventures rather than to adopt e-commerce, loses information about its core competencies, significantly improve profits sales and importance of operating business by using internet in the future. Relative advantage is one of the most frequently used innovation characteristics in adoption research.

5. Complexity

Tests on the complexity factors carried out by various studies in several places, including Singapore [9] and Indonesia. Research conducted in Indonesia conducted in Yogyakarta [4] and Pekanbaru [11]. Complexity is also one of the characteristics of the adoption of innovations Rogers that are often studied.

6. Trialability

The perceived high cost of implementing e-commerce is a determinant. Availability and awareness of grants available are relevant to this factor. When grants are given, businesses are able to reduce high start up costs. Pain associated with failed initiatives is also lessened. In short, grants will enhance the trialability of the innovation [9].

This study proposed the following hypotheses:

H1- H6: There were no significant difference between adopters and non-adopters SMEs in terms of each factor that was described above.

5.2 Sampling Method

The surveys were distributed directly to the respondent, and then it filled in with guidance, so that respondent could understand the questions clearly. The surveys were targeted so that the respondent is most likely the owner/CEO of the business, which in most SMEs is the same person.

Proportional stratified random sampling and quota sampling was used as method based on the limitation defined, the SMEs are located in Jakarta, Bogor, Depok, Tangerang and Bekasi (Indonesia).

Composition of data sample between adopters and non-adopters sought to be collected in a balanced way so that the results would be more accurate.

Number of data sampling for each sector shown as follows:

Table 1. Number of Samples for Each Sector

Sector	Number of SMEs	Percentages	Number of Samples
Craft Industry	76	38%	19
Garmen industry	54	27%	14
Other	70	35%	17
Total	**200**	**100%**	**50**

Data were collected from the respondent, and then were processed using Mann-Whitney U-Test. This method of processing data was choosen because:

1. The dependent variable of adopters and non-adopters could be regarded as two independent sets of data.
2. The data can not be assumed were obtained from the normal distribution population.

5.3 Questionnaire Design

The structure of the questionnaire was divided into three major sections as follows:

1. Company and business owners profile
2. Use of information technology
3. E-commerce adoption factors

The first section of the survey relates to the SMEs' company and business owners profile. Data was taken to classify respondents based on several categories. Although it were not included in the analysis of e-commerce adoption factors, it were very useful supporting the final conclusions. The section was continued with the use of information technology to measure whether these SMEs have adopted e-commerce technology or not adopted yet. The final section of the survey gathered main informations for measuring the following factors:

1. Resources and readiness of the organization, with the following indicators:
   a. Financial resources
   b. Technology resources
   c. Readiness of existing technology
   d. Human resources
   e. Top level management understanding of technology
   f. Governance procedures within the organization
2. Management support, with the following indicators:
   a. Development of vision and strategy for e-commerce
   b. Communications support
   c. Policies and organizational structures
3. Environmental Pressure, with the following indicators:
   a. Encouragement from similar business success in implementing e-commerce
   b. Encouragement to compete within a similar sector
   c. Encouragement from stakeholders (customers, suppliers, investors, media, business associates and employees)
   d. Support services from technology vendors and government regulations
4. Relative advantage, with the following indicators:
   a. Lower business costs
   b. Speed up the completion of work
   c. Reach a broader market
   d. Preference to upgrade other business ventures rather than to adopt e-commerce
   e. Loses information about its core competencies
   f. Significantly improve profits sales
   g. Importance of operating business by using internet in the future
5. Complexity, with the following indicators:
   a. Technical knowledge
   b. Safety factor
6. Trialability, with the following indicators:
   a. The initial cost for implementation of e-commerce
   b. Availability of subsidies or loan funds from investors, governments, and other parties
   c. Access opportunities to low-cost media such as mailing lists, forums and other internet media

The third section of the questionnaire using closed questions (closed-ended questionnaire) and using a 1-7 Likert scale.

6. FINDINGS AND DISCUSSION

Once the questionnaire data had been collected, I checked whether the data was valid for the next processing step or not. I used Cronbach's alpha to assess the reliability or internal consistency of each multi-item measurement in the interval scale.

1. Validity Test

The results of Pearson product moment count for each item in the question exceeded the r table 0.279. Therefore it concluded that the whole question in the questionnaire were positively correlated with total score, so that everything could be included in the subsequent analysis.

2. Reliability test

The reliability test results showed alpha values as in the following table:

Table 2. The results of Reliability Test

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Cronbach's alpha	0,805	0,863	0,808	0,783	0,844	0,840

The results obtained in Table 2 above were compared with the r table 0.279. All alpha values were above the value of that r table, so it concluded that the question in the test instrument is all reliable to be processed.

6.1 General Description of Respondents

Respondents in this study were classified in several aspects, they were:

1. Age of respondents

The greatest proportion (38%) of respondents was in the productive age of 30-40 years. Most of the respondents are the owners of SMEs who had great power for taking a decision to the company.

2. Education level of respondents

Above 50% respondent, the majority of SMEs' owners had been well educated. The education level was above high school and degree. This result was related to the internet technology open minded.

3. Number of employees

The survey result shown that 64% of respondents had fewer than 10 employees, this related to the allocation of human resources in the organizations.

4. Year of business was founded

The results shown that the majority of 52% respondents were established a business between the years 2005 and 2010. The most of the age of SMEs respondents was quite young (below five years).

5. Estimated number of annual turnover

Respondents could be classified into three categories as follows:

a. Below 100 million rupiahs (36% respondent)
b. Between 100-300 million rupiahs (34% respondent)
c. Above 300 million rupiahs (30% respondent)

6. Use of Information Technology

The survey results showed that the majority of business owners involved in the decision of the use of computers and information technology in the company. The next question was about the frequency of computer use and Information Technology by SMEs' owners. From the survey, there were 46% of respondents replied that the owners of SMEs rarely use computers and information technology in daily activities. How about the SMEs' access to the computers and the internet? The majority of respondents, 72% had used computer technology and 54% were internet users, while the rest 32% were users of other technologies (e.g. mobile phones). What about the use of email? The survey showed that the 68% of respondents had been using email in their business. This was regardless of whether the email was used as a communication channel for business e-commerce or merely just to filled the requirements of the SMEs membership SMESCO Trading Board.

6.2 SMEs Adoption Factors

From the 50 respondents had been studied, there were 22 respondents (44%) have adopted e-commerce and 28 respondents (56%) have not adopted e-commerce.

The result of Mann-Whitney U-Test showed that value of *Asymp. Sig (2-tailed)* of all factors seemed bellow $\alpha = 0.05$. Therefore, All of hypothesses had been rejected. The one factor, resources and organizational readiness, have shown 0.00 as value of *Asymp. Sig (2-tailed)*. From that result, I concluded that resources and organizational readiness were very significant to the SMEs' e-commerce adoption. Summary of the significance level of e-commerce adoption factors are shown in Table 3 below:

Table 3. Summary of Hypotheses Testing Results

Category	Factors	*Asymp. Sig (2-tailed)*	Significancy Factors
Organizational Context	Resources and organizational readiness	,000	Very significant
	Management support	,002	Significant
Environmental Context	Environmental pressure	,001	Significant
Innovation characteristics Context	Relative advantage	,001	Significant
	Complexity	,010	Significant
	Trialability	,007	Significant

In this section the research problem had been answered. There were significant difference between adopters and non-adopters, so the model that was developed could be used to determined adoption of ecommerce by SMEs especially in Indonesia.

7. CONCLUSION

The organization readiness was the most important thing that needs to be improved in SMEs for increasing the level of adoption of e-commerce. As an attempt to explore this factor, respondents have been interviewed. I found the fact that most respondents got difficulties to adopt e-commerce due to lack of allocation of human resources who can manage business lines of e-commerce. SMEs had limited person to run their business. Mostly the owner was the only one who runs the business. Another reason was lack of time and skill to learn new

technology, and also lack of financial capability to buy new hardware. In fact, most of them did not have specific computer for regular business use. This organization readiness is closely related to the second factor, management support to e-commerce business. If the top management has a great commitment to e-commerce adoption, the allocation of resources will not become a barrier anymore.

Moreover, environmental pressure was also became a determinant for the adoption of e-commerce, which means SMEs that did not adopted e-commerce had unfavorable environment of doing business via the internet so that they were not compelled to adopt it.

Another factor is relative advantages. The non-adopters had not understood the great benefits that may be acquired by doing the business operation via internet. Most of SMEs found that the traditional lines of business were more profitable than those e-commerce business lines. The non-adopters thought that e-commerce business had a high complexity, in the context of safety and technical issues. Therefore the encouragement and adequate training for SMEs was needed to enhance the understanding of safety and technical issues.

8. FURTHER WORK

Based on several findings in this research, suggestions are given for further work as follows:

1. Increasing the number of samples would raise the research accuracy level, so it can be used as to generalize the results of research, close to the reality.

2. To anticipate the low response rate, further research should collaborate with related institutions, so that respondents are concerned to fill in the questionnaire. As an example, the research could probably be included in the census survey of SMEs.

3. Future studies may use a field worker to spread and then return the questionnaire, so that the samples can be randomly taken and representative for overall industrial sector. It could satisfy the normal distribution of a population.

4. Qualitative assessments as additional information can enrich the research study. It can be done by extending an open field in the questionnaire and adding longer period of research so that the respondent could be more intensively interviewed.

9. REFERENCES

[1] M. J. Hafsah, "Upaya Pengembangan Usaha Kecil Dan Menengah (UKM)," *Infokop,* no. 25, 2004.

[2] D. Chaffey, "E-business and E-commerce Management," in *E-business and E-commerce Management*, Edinburgh, Prentice Hall, 2007, pp. 8-9.

[3] T. Oswari, S. Suhendra and A. Harmoni, "Model Perilaku Penerimaan Teknologi Informasi: Pengaruh Variabel Prediktor, Moderating Effect, Dampak Penggunaan Teknologi Informasi Terhadap Produktivitas dan Kinerja Usaha Kecil," *Seminar Ilmiah Nasional Komputasi dan Sistem Intelijen (KOMMIT 2008),* 2008.

[4] F. Wahid and L. Iswari, "Adopsi Teknologi Informasi Oleh Usaha Kecil Dan Menengah Di Indonesia," *Seminar Nasional Aplikasi Teknologi Informasi (SNATI),* pp. J75-J79, 2007.

[5] E. M. Rogers and K. L. Scott, "The Diffusion of Innovations Model and Outreach from the National Network of Libraries of Medicine to Native American Communities," 1997. [Online]. Available: http://nnlm.gov/archive/pnr/eval/rogers.html. [Accessed 20 April 2010].

[6] S. Kurnia, B. Alzougool, M. Ali and S. M. Alhashmi, "Adoption of Electronic Commerce Technologies by SMEs in Malaysia," *Proceedings of the 42nd Hawaii International Conference on System Sciences,* 2009.

[7] L. Sparling, M. Toleman and A. Cater-Steel, "SME Adoption of e-Commerce in the Central Okanagan Region of Canada," *18th Australasian Conference on Information Systems SME Adoption of e-Commerce,* pp. 1047-1059, 2007.

[8] N. Al-Qirim, "The Adoption of eCommerce Communications and Applications Technologies in Small Businesses in New Zealand," *IEEE International Conference on Computer Systems and Applications,* pp. 894 - 897, 2006.

[9] J. Kendall, L. L. Tung, K. H. Chua, C. H. D. Ng and S. M. Tan, "Electronic Commerce Adoption by SMEs in Singapore," *Proceedings of the 34th Hawaii International Conference on System Sciences,* 2001.

[10] C. Lertwongsatien and N. Wongpinunwatana, "E-commerce adoption in Thailand: An Empirical Study of Small and Medium Enterprises (SMEs)," in *Journal of Global Information Technology Management*, Georgia, 2003.

[11] B. L. Hakim, "Studi priliminer pemanfaatan internet e-commerce pada usaha kecil menengah (UKM) di Pekan Baru Propinsi Riau," Pascasarjana MTI - Fasilkom UI, Jakarta, 2004.

[12] H. N. Siddiqi, "Adopsi E-Commerce untuk Usaha Kecil dan Menengah studi kasus Bank Perkreditan Rakyat di Kota Depok," Pascasarjana MTI - Fasilkom UI, Jakarta, 2007.

Collision Risk Modeling Using Monte-Carlo Simulation

Moeljono Widjaja
Agency for The Assessment and Application of Technology
BPPT Tower 2, Level 18
Jl. M.H. Thamrin No. 8
Jakarta 10340, Indonesia
moeljono.widjaja@bppt.go.id

ABSTRACT

This paper presents an improved method in estimating air traffic collision risk based on a well-known Reich's Collision Risk Model. The method focused on lateral separation in line with the introduction of RNP (Required Navigation Performance) / RNAV (Area Navigation) routes in many major airspaces. Implementation of RNP/RNAV routes in some airspace was intended to increase efficiency and capacity of airspace in order to cope with continual growth of air traffic. The method proposed a decomposition of the Reich's model into two parts: deterministic and stochastic. The stochastic components were solved using Monte-Carlo Simulation while the deterministic components were calculated from Traffic Sample Data (TSD). Route A576 in Indonesian Flight Information Region (FIR) was used as the object of study in implementing the proposed method.

Keywords

Air Traffic Management, Reich's collision risk model, Monte-Carlo Simulation

1. INTRODUCTION

Air traffic volume in Indonesian airspace has increased dramatically since the introduction of de-regulation in the airline industry in 2000. This phenomenon has prompted the regulator, Directorate General of Civil Aviation (DGCA), to take an immediate action by implementing PBN (Performance-Based Navigation) in some of international routes in order to increase the capacity of Indonesian airspace while keep maintaining the safety level such that the estimated total risk is below the Target Level of Safety (TLS) of 5 x 10^{-8} fatal accidence per flight hour. For example, creating a new track between any two existing adjacent tracks will double the capacity of the airspace. However, safety level must be maintained by implementing certain measures such as RNP (Required Navigation Performance)/RNAV (Area Navigation) specifications on the designated routes. RNP/RNAV10 is defined that any aircraft flying in this route must remain within five nautical miles of the assigned flightpath at least 95% of the total flying time.

The implementation of PBN should follow the guideline published by International Civil Aviation Organization (ICAO) which is Document 9613 [5]. It also requires a regular assessment on the safety level by a monitoring agency appointed by ICAO. ICAO has recommended a classical Reich's Collision Risk Model to assess the safety level of implementation RNP as explained in ICAO Document 9689 [4].

This paper discussed and proposed an improved method in the model by introducing Monte-Carlo Simulation to estimate the probabilistic parameters of the model. Firstly, a lateral separation model which was the focus of this paper was defined. Deterministic parameters of this model were evaluated and discussed. Secondly, stochastic components of the model (the probability of lateral and vertical overlaps) were estimated by Monte-Carlo Simulation. Thirdly, the model was tested on Traffic Sample Data collected for Route A576 and simulated for a few scenarios such as varying system occupancy. Finally, it gave some recommendation and conclusions. It is worth noted that the proposed method could also be further explored and implemented as a safety assessment tool as proposed in [7].

2. METHODOLOGY

This section discussed a methodology in calculating the estimated risk of collision between aircrafts separated laterally by a standard separation minima. They included definition of the lateral collision risk model, characteristics of the object of study, and methods in estimating the probability of lateral/vertical overlap using Monte-Carlo Simulation.

2.1 Lateral Collision Risk Model

Collision Risk Model is a model being used to estimate risk of collision between aircrafts flying in their designated paths due to their flying errors. The model was originally developed by P.G. Reich as explained in [8, 9, 10]. The model took into account the movement of aircrafts in all three directions: longitudinal, lateral and vertical. Target Level of Safety is maintained by applying separation standards in the three directions.

This paper only focused on a lateral separation between aircrafts in adjacent routes which would have impacts on the implementation of RNP/RNAV routes. Reich's lateral collision risk model is described in the following Equation:

$$C = P_y(S_y)P_z(0)\frac{\lambda_x}{S_x}[E_s\left[\frac{|\bar{\dot{x}}_s|}{2\lambda_x}+\frac{|\bar{\dot{y}}|}{2\lambda_y}+\frac{|\bar{\dot{z}}|}{2\lambda_z}\right]+ E_o\left[\frac{|\bar{\dot{x}}_0|}{2\lambda_x}+\frac{|\bar{\dot{y}}|}{2\lambda_y}+\frac{|\bar{\dot{z}}|}{2\lambda_z}\right]] \quad (1)$$

C the expected number of fatal accidents per aircraft flying hour.

$P_y(S_y)$ lateral overlap probability, i.e. the likelihood that any two aircrafts which have been assigned the correct lateral separation are in fact not separated laterally.

$P_z(0)$ probability that two aircrafts assigned to the same flight level are in vertical overlap.

$|\bar{\dot{x}}|$ average along-track component of the relative velocity of two aircrafts which collide due to loss of longitudinal separation.

$|\bar{\dot{y}}|$ average relative cross-track speed for two aircrafts assigned to the same track.

$|\bar{\dot{z}}|$ average relative vertical speed for two aircrafts assigned to the same level.

λ_x average length of aircraft using the airspace.

λ_y average wing-span of aircraft using the airspace.

λ_z average vertical dimension of aircraft using the airspace.

S separation standard in use. S_x is the longitudinal separation standard, and S_y is the lateral separation standard.

E the systems occupancy and provides a measures of the traffic density. The system occupancy (E_s and E_o where "s" represents same direction traffic and "o" represents opposite direction traffic) is estimated from flight plan data for a set of sample days throughout the year.

Equation 1 was analyzed based on the state of each variable: independent/dependent and time varying/invariant. The independent variables were clearly S_y and S_x. Variables on Equation 1 can also be categorized into two groups: time-dependent variables and time-independent variables. Time-dependent variables were $P_y(S_y)$, E_s and E_o. Other variables were relatively constant with time.

Figure 1 shows two adjacent tracks (A and B) with opposite traffic direction where the tracks were separated by a lateral separation standard S_y and a longitudinal separation standard S_x. In this scenario, Equation 1 can be simplified to the following Equation:

$$C = P_y(S_y)P_z(0)\frac{\lambda_x}{S_x}E_o\left[\frac{|\bar{\dot{x}}_0|}{2\lambda_x}+\frac{|\bar{\dot{y}}|}{2\lambda_y}+\frac{|\bar{\dot{z}}|}{2\lambda_z}\right] \quad (2)$$

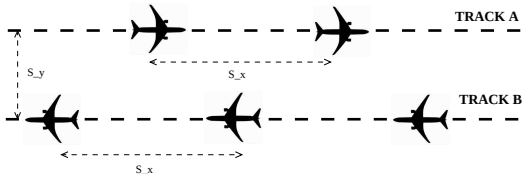


Figure 1: Parallel tracks with opposite direction traffic.

2.2 The Object of Study: Route A576

Indonesian airspace was divided into two Flight Information Regions (FIR): Jakarta (Western area) and Ujung Pandang (Eastern area). The Route A576 consisted of ten way points and spanned across the two FIRS. Table 1 shows the Waypoint Position Reporting (WPR) Points of the A576 Route and the distance between the WPR point and the next WPR point. The average distance between fixes in Route A576 was 114.7 NM. The total distance of Route A576 from SANOS to ATMAP was 1032 NM.

Table 1: Waypoint Position Reporting (WPR) Points of the A576 Route

WPR Point	**Latitude** (°S)	**Longitude** (°E)	**Distance to Next WPR Point** (NM)
SANOS	00 42 00.00	105 19 00.00	172
APARI	02 02 36.50	107 50 25.38	62.7
APAGA	02 46 57.63	108 34 51.23	30
AKULA	03 07 11.56	108 57 03.33	80
SABIL	04 01 08.65	109 56 18.70	68.9
SUMBU	04 45 18.00	110 49 24.00	139
SIPUT	06 17 00.00	112 35 00.00	213
BLI	08 45 02.05	115 09 47.90	150
ILDAM	10 35 31.61	116 53 04.43	117
ATMAP	12 00 00.00	118 15 18.00	

Some parameters of the model were obtained from Traffic Sample Data (TSD) in Ujung Pandang FIR (Flight Information Region) for the full month of December 2010. The average dimensions of the aircrafts flying in Ujung Pandang FIR were $\lambda_x = 0.031783$ NM, $\lambda_y = 0.028710$ NM, and $\lambda_z = 0.008958$ NM. The average cruising speed of flights flying in A576 route was 468 knots. This speed was calculated from the TSD which included all flights flying above 245 FL. The normal speed of aircrafts was assumed to be within the range of 250 and 650 knots; any speed outside this range was considered as outliers. Given the total distance of route A576 (SANOS to ATMAP) and the average cruising speed of aircrafts flying on route A576, the average flying hour for an aircraft on route A576 was found to be 2.2 hours.

Proximity time is calculated from the aggregate time where two aircrafts are in proximate pairs. Proximate pairs occur when two aircrafts in adjacent tracks are within a distance of the longitudinal separation standard/minimum (i.e., S_x). For example, if two aircrafts were flying on adjacent tracks and in opposite direction, then the total proximity time was $\frac{2\times S_x}{|\bar{\dot{x}}_0|}$ where $S_x = 120$ NM and $|\bar{\dot{x}}_0|$ was twice of the average cruising speed ($\bar{V} = 468$ knots). Therefore, the total proximity time was 0.256 hours.

In [9], the lateral occupancy for opposite direction traffic E_o

is defined as:

$$E_o = \frac{2 \times T_y}{H} \tag{3}$$

where

- T_y = the aggregate of time spent when all pairs were in the configuration of lateral proximity
- H = the total flying hours of all aircrafts during the period considered.

Given above conditions, for a pair of aircrafts flying on route A576 in opposite direction traffic, the lateral occupancy was 0.116.

The conventional route A576 was replaced by the new RNAV route M635 on 9 February 2012 as published in [2]. Fixes of the new route M635 are SURGA, RUSMA, SAMSU, MASRI, RAFIS, TAVIP, SUMDI, RAMPY, UDONO, BLI, ATMAP. The new route reduced the lateral separation minima from 60 NM to 50 NM, and required the standard deviation of lateral track errors to be less than 8.7 km (4.7 NM).

The vertical deviation error was defined by the ASE (Altimetry System Error) model. Since there was not any data or information about ASE model for aircrafts flying in Indonesian airspace, we used the ASE model from the ARINC report [6]. The mean and standard deviation of probabilistic density function of the vertical deviation were -4.38 feet (-0.000720 NM) and 44.14 feet (0.007264 NM) respectively.

The average relative vertical speed for two aircraft assigned to the same level was assumed to be 1.5 knots and the average relative cross-track speed was 13 knots. These were commonly used values as found in the RVSM (Reduced Vertical Separation Minima) Safety Assessment Report [1].

2.3 The Probability of Lateral Overlap

The probability of lateral overlap is defined as the likelihood that any two aircrafts which have been assigned the correct lateral separation are in fact not separated laterally. The probability distribution function which best represented the probability at the tail-end was Laplace distribution. Laplace Probability Density Function (pdf) is given as follows:

$$f(y) = \frac{1}{\sqrt{2}\sigma} \exp^{-\frac{\sqrt{2}|y-\mu|}{\sigma}} \tag{4}$$

where

- μ = mean
- σ = standard deviation / scale

The corresponding probability distribution graph is shown in Figure 2. The intersection of the two identical Laplace probability distribution functions where the means were separated by S_y was at $\frac{S_y}{2}$.

The probability of lateral overlap $P_y(S_y)$ is equal to the shaded area of Laplace probability distribution graph as shown in Figure 2.

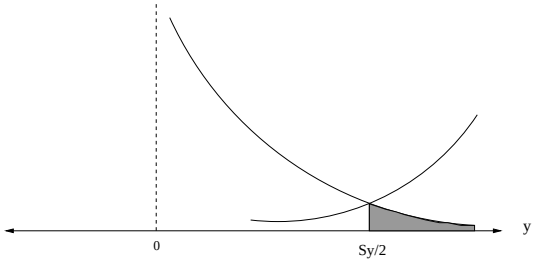


Figure 2: Laplace probability distribution graph.

$$\begin{aligned} P_y(S_y) &= \text{shaded area} && (5)\\ &= \int_{\frac{S_y}{2}}^{\infty} f(y)\,dy && (6)\\ &= \left[-\frac{\sigma}{\sqrt{2}}\frac{1}{\sqrt{2}\sigma}\exp^{\frac{\sqrt{2}|y-\mu|}{\sigma}}\right]_{\frac{S_y}{2}}^{\infty} && (7)\\ &= \frac{1}{2}\exp^{\frac{\sqrt{2}\left|\frac{S_y}{2}-\mu\right|}{\sigma}} && (8) \end{aligned}$$

For RNP/RNAV10 routes, the lateral error deviation of aircrafts flying in these routes can be described by the following equation:

$$\int_{-5}^{5} f(y)\,dy = 0.95 \tag{9}$$

Basically, Equation 9 states that lateral deviation of the aircraft must be bounded by ± five nautical miles from the assigned track at least 95% of the total flying hours. In order to meet the above condition, the standard deviation of the Laplace distribution function can calculated by solving Equations 4 and 9. The calculated value of σ was found to be 2.36 NM. Consequently, the probability of lateral overlap in RNP/RNAV10 routes with lateral separation $S_y = 50$ NM or $P_y(50)$ was found to be 1.5625 x 10^{-7}.

Alternatively, the probability of lateral overlap can be estimated by Monte-Carlo Simulation. Five simple steps to run Monte Carlo Simulation as explained in [11]:

1. Define a parametric model, $y = f(x_1, x_2, ..., x_q)$.
2. Generate a set of random inputs, $x_{i1}, x_{i2}, ..., x_{iq}$.
3. Evaluate the model and save the results as y_i.
4. Repeat step 1 and 2 for $i = 1$ to n.
5. Analyze the results using histogram, summary statistics, etc.

The Monte Carlo Simulation module to calculate $P_y(S_y)$ was written in Octave [3]. The random number generator used in this simulation was Laplace distribution function.

The parametric model to calculate $P_y(S_y)$ is given by:

$$D_y = S_y - (\lambda_y + |y_{AA'} + y_{BB'}|) \quad (10)$$

where

S_y is the standard lateral separation minimum

λ_y is the average wingspan of aircrafts

$y_{AA'}$ is the lateral deviation error of aircraft A

$y_{BB'}$ is the lateral deviation error of aircraft B

If D_y is less than or equal to zero, then a lateral overlap occurs. Figure 3 illustrated the lateral overlap between two aircrafts A and B which flying on adjacent tracks and at the same flight level.

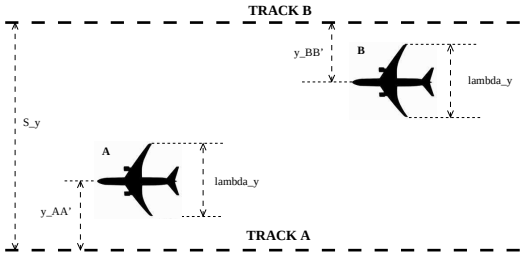


Figure 3: Lateral overlap.

Figure 4 shows the histograms which were the results of running Monte-Carlo simulation under the following scenario:

- Random number generator was based on Laplace distribution with parameters ($\mu = 0$ dan $\sigma = 2.36$ NM).
- Lateral separation minima $S_y = 50$ NM.
- Number of samples $n = 10000000$.

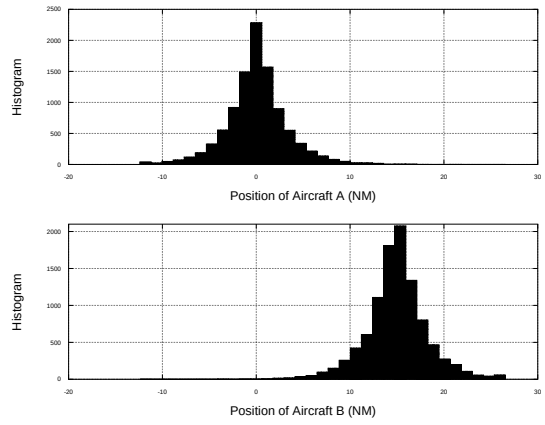


Figure 4: Histograms of two aircrafts A and B in adjacent RNP/RNAV10 routes with lateral separation S_y = 50 NM.

2.4 The Probability of Vertical Overlap

Similarly, the probability of vertical overlap can be estimated using Monte-Carlo Simulation method as explained earlier.

The parametric model to calculate the probability of vertical overlap $P_z(0)$ is given by:

$$D_z = |z_{AA'} - z_{BB'}| - \lambda_z \quad (11)$$

where

λ_z is the average height of aircrafts

$z_{AA'}$ is the vertical deviation error of aircraft A

$z_{BB'}$ is the vertical deviation error of aircraft B

If D_z is less than or equal to zero, then a vertical overlap occurs. Figure 5 illustrated the vertical overlap of two aircrafts A and B which flying on adjacent tracks and at the same flight level.

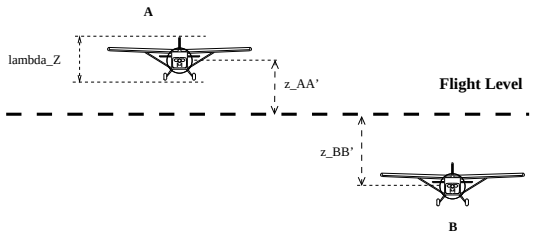


Figure 5: Vertical overlap.

The module of Monte-Carlo Simulation for estimating the value of $P_z(0)$ was also written in Octave [3]. The random number generator used in this simulation was Normal Probability Distribution function.

The results of the simulation is shown in Figure 6. Two histograms (subplot 1 and 2 in Figure 6) shows the vertical error distribution from aircraft A and B. Subplot 3 shows the distribution of vertical distance between aircraft A and B. The vertical deviation distribution was assumed as a normal distribution with mean = -0.000720 NM and standard deviation = 0.007264 NM. Given the average height of the aircrafts $\lambda_z = 0.0099$ NM, the probability of vertical overlap $P_z(0)$ was found to be 0.61715 (based on the result of Monte-Carlo Simulation with $n = 10000000$).

3. RESULTS AND DISCUSSION

This section presents the results of lateral collision risk estimation with varying system occupancy on A576 route. Table 2 shows parameters of lateral collision risk model for the A576 route. The values of these parameters were already discussed in the previous sections.

Figure 7 shows the calculated risk of collision as a function of system occupancy and lateral separation: increasing the system occupancy will increase the risk while decreasing the lateral separation will also increase the risk. The lateral separation minima was determined by RNP/RNAV specifications of the routes so that it could not be furthered reduced

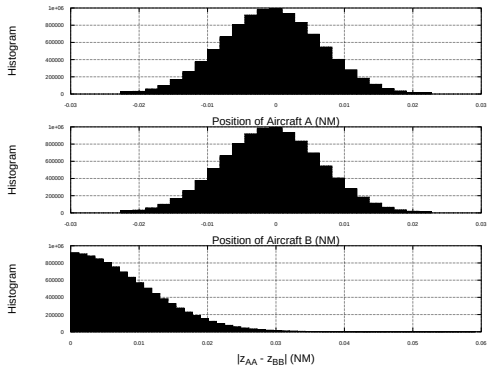


Figure 6: Simulation results of $P_z(0)$.

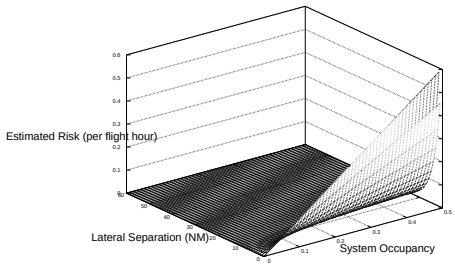


Figure 7: Estimated risk as a function of system occupancy and lateral separation.

Table 2: Parameters of Lateral Collision Risk Model for A576 Route

Parameter	Estimated Value
S_y	50 NM
S_x	120 NM (15 mins)
$P_z(0)$	0.61715
λ_x	0.031783 NM
λ_y	0.028710 NM
λ_z	0.008958 NM
$\lvert\dot{x}_0\rvert$	936 kt
$\lvert\dot{y}\rvert$	13 kt
$\lvert\dot{z}\rvert$	1.5kt

or optimized. In contrast, system occupancy is an important component that has significant impact in determining the collision risk. The system occupancy was determined by many factors such route designs and aircraft scheduling (traffic distribution and density). Optimization of air traffic flow management can help in keeping the system occupancy at relatively low level so that minimizing the risk of collision.

4. CONCLUSIONS

This paper has presented a technique in estimating the collision risk for aircrafts flying in parallel tracks in opposite traffic direction with a given lateral separation standard. The introduction of Monte-Carlo simulation gave flexibility in modeling aircraft movements due to flying errors. However, Monte-Carlo simulation required extensive usage of computing power and large memory that can limit the complexity of the model in the simulation. A more practical and efficient technique is required in order to run the simulation for a large and complex model. Additionally, theoretical and practical techniques should be combined in order to have an accurate and yet practical technique in estimating the risk using available data which are often very limited.

A few conclusions and recommendations from this paper can be drawn as follows:

- ASE model for aircrafts flying in Indonesian airspace is required to have more accurate estimation of the vertical overlap probability.
- Impact of traffic distribution on estimated risk should be analyzed.
- A further study is required in managing system occupancy by designing parallel tracks in same traffic direction but with efficient scheduling of airplanes in order to increase traffic density but keeping the system occupancy relatively low.
- A further work is also needed to integrate the lateral collision risk model to a complete model which includes impacts of vertical and longitudinal separations.

Acknowledgments

The research was fully supported by CNS/ATM Project, Pusat Teknologi Informasi dan Komunikasi (PTIK), Badan Pengkajian dan Penerapan Teknologi (BPPT), Jakarta, Fiscal Year 2010-2012.

5. REFERENCES

[1] AAMA. Airspace safety review of the rvsm implementation in indonesian airspace march 2010 to february 2011. RVSM Safety Assessment Report, March 2011.

[2] DGCA. Restructuring of ats routes in the jakarta fir and ujung pandang fir. AIP Supplement 11, December 2011.

[3] J. W. Eaton, D. Bateman, and S. Hauberg. *GNU Octave Manual: a high-level interactive language for numerical computations*. SoHoBooks, 3 edition, 2007.

[4] ICAO. *Manual on Airspace Planning Methodology for the Determination of Separation Minima*. International Civil Aviation Organization, 1 edition, 1998. Doc 9689-AN/953.

[5] ICAO. *Performance-Based Navigation (PBN) Manual*. International Civil Aviation Organization, January 2008. Doc 9613.

[6] G. Moek, E. Lutz, and W. Mosberg. Risk assessment of rnp10 and rvsm in the south atlantic flight identification regions. Technical report, ARINC Incorporated, May 2001.

[7] I. Premiramanda and M. Widjaja. Collision risk simulator: a safety assessment tool. In *Proceedings of 10th WSEAS International Conference on*

Computational Intelligence and Information Security, pages 143–149, Jakarta, December 2011.

[8] P. G. Reich. Analysis of long-range air traffic systems: Separation standards i. *The Journal of Navigation*, 19(01):88–98, 1966.

[9] P. G. Reich. Analysis of long-range air traffic systems: Separation standards ii. *The Journal of Navigation*, 19(02):169–186, 1966.

[10] P. G. Reich. Analysis of long-range air traffic systems: Separation standards iii. *The Journal of Navigation*, 19(03):331–347, 1966.

[11] J. Wittwer. "monte carlo simulation basics" from vertex42.com. http://www.vertex42.com/ExcelArticles/mc/MonteCarloSimulation.html, June 2004.

Designing and Developing Petra Christian University Learning Management System

Arlinah I.Rahardjo
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
+62312983456
arlinah@petra.ac.id

Andreas Handojo
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
+62312983455
handojo@petra.ac.id

Jeremy Martinus Karyadi
Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
jayree87@gmail.com

ABSTRACT
Teaching and Learning process can be conducted onsite and online. The online process is known as a method of learning called Learning Management System (LMS). This online system is generally meant to enrich the traditional teaching-learning process. Petra Christian University (PCU) has been implementing two types of LMS, Moodle-based LMS called E-course and PCU Camel. Those two LMS are not yet integrated into the PCU MIS. This dis-integration creates duplication efforts in data gathering process. An integration application system is needed to be developed in term of user management, course management as well as grading system to create an effective and efficient use of LMS. Based on the black box testing and user acceptance testing for both types of users of the system administration site and course administration site, Lentera LMS, the name of the new LMS, proved to be able to meet the user requirements.

Keywords
Learning Management System, Moodle, Integration System, Lentera, Petra Christian University

1. INTRODUCTION
Interactions among students and teachers are of vital importance in teaching-learning processes. Various methods have been designed to facilitate those interactions, onsite or online. In the meantime teaching learning process is generally conducted onsite in a traditional way, especially in Indonesia. These onsite processes have limitation in term of time and space to allow maximum interactions.

The progress of information technology gave birth to new teaching-learning methods, using online system. This new online system reduces the dependency of learning in term of time and space barrier and offers flexible and maximum interactions among teachers and students. This method also opens new forms of learning enrichment. The use of web for a learning application is called Learning Management System (LMS).

For many years, Petra Christian University (PCU) has been experimenting and developing several types of LMS to enrich the learning experiences of the students and to allow more flexible interactions among students and teachers. In the recent years, PCU has been implementing two types of LMS. The first LMS, called E-course, is an open source, Moodle-based LMS. The second one is called PCU-Camel (Petra Christian University – Computer Aided Mechanical Engineering Learning Environment), developed as the final work of an Informatics Students of PCU. PCU-Camel is an in-house application using linux. Each LMS has its own strengths and weakenesses. However, both applications have common problems. They are not integrated into the existing university administration system, dealing with the management of the data on students, staffs, courses or grading sstem.

The teachers have to contact the LMS administrator to allow their courses to be accessed by their students. The teachers also have to register to the system by supplying their personal data, the course data, and the identity of the students registering in their classes, whereas those data have already existed in the University administration online system. This duplication also happens in the grading system. These duplications of data create ineffectiveness and inefficiency in the use of the LMS. The grades of the students recorded in the two LMS also can not be transferred to the university grading system developed by PCU MIS, which has its own system with different database platform. Those problems prevent the two LMS to be widely used by PCU teachers.

To solve those above problems, an LMS system integrated into the existing university system is needed to be developed. The records of the teachers, students and courses need to be automatically activated as the teachers are assigned to teach in certain courses and as the students registered to the courses every new semester. The LMS system will also provide a facility to transfer the students's grades to the university administration system to prevent any data duplication. It is expected the integrated new LMS system will offer more effective and efficient system to both teachers and students.

2. THEORETICAL BACKGROUND
2.1 Learning Management System
Learning Management System (LMS) or *Course Management System* (CMS) is a web-based application to enrich teaching-learning process by making use of internet without replacing the roles of teachers [1].
A web-based LMS has several strengths to allow the effectiveness of teaching learning process, as follows:

- **Personalized:** having the capacity to adjust the program based on the needs of the institution, departments or students individually.
- **Interactive**: providing facilities for students to collaborate with other students as well as teachers.
- **Just-in-time**: giving the opportunities to students to participate in the unscheduled learning activities.
- **Current** : providing updated course materials
- **User-centric**: focussing on the needs of the students.

An LMS runs on a web server and is accessed by using a web browser application. The basic feature of an LMS is a learning aid to create web-based courses and the access rights management for participants of certain courses. An LMS also provides various features to let the course be more effective such as uploading and

sharing course materials, online discussion, chatting, quizzes, surveys, reviewing students' assignments and saving students' grades [1].

In an LMS, there is generally an administration system managed by an administrator. An administrator has the highest access rights in the LMS. The access right will allow each user to have different limitation in accessing the features in the LMS [2].

2.2 Moodle

Moodle is an open source Course Management System (CMS) implemented by the universities, learning communities, K-12 schools, business enteprises or even individual intructors to include the use of web technology to the teaching-learning process[1]. Moodle is an acronim of Modular Object-Oriented Dynamic Learning Environment, developed by Martin Dougiamas. He is a computer expert as well as an educator who has spent many hours to support a learning management system in an Australian university. Acoording to Martin, a learning system will be a better system if the development is based on the learning theory, not on the technical aspect only [1].

Moodle has several strong points compared to other CMS as follows [1]:

- **Free and Open Source:**
  The idea of open source means the users having access to the source code of software. This characteristic is important as this software is inter-related to an academic community who appreciate freedom and knowledge sharing. Moodle can be downloaded freely. Users can develop new features, fix errors, increase the capability of the application and undershand how other people solve the problems in the programming process. Besides, unlike other expensive system requiring licence fees and maintenance contract, Moodle does not need any fees to download and install the application in any server.
- **Educational Philosophy**
  Martin Dougiamas' background in the area of education directed him to adopt the development process Moodle from the social aspect (*social constructionism*) as the nucleus theory behind Moodle. If most CMS is developed as a tool-centered application, Moodle is developed as learned-centered application.
- **Social contructionism** is based on the idea that a learner can understand best if he/she is participating in the social process to contribute something worthwhile to other persons. The social process indicates that the learning process happens in a team. From this perspective, learning process is a process of exhanging ideas to search for the meanings of something worthwhile to be shared. This process is the process of building knowledge.
- **Community**
  Moodle has quite a big number of active communities, whose members are using the system, develop new features and increase the quality of the CMS. This community is priceless for the success of a system. This community helps offering solutions for problems encounted by the users of the system. At the same time, the ones who develop and the ones who use the system collaborate to guarantee the quality of the system, add modules and features and give ideas for development.

3. SYSTEM DESIGN

The new LMS system for Petra Christian University is named Lentera (Learning through Petra). It is designed and developed using Moddle, as Moodle provides administration features that have not been explored much at E-course LMS, while the problems found are dealing more with administration system. The administration features at Moodle are more adaptable to be used to solve the above-mentioned problems at PCU as well as to accommodate future developments. PCU-Camel is a customized system not flexibled enough to be developed further to accommodate the features needed to be integrated into PCU MIS.

The development of the Lentera LMS is focused on the administration system of LMS. The features include the the integration of students', teachers', and course' data into those data available at PCU MIS. Those features include User Management, Course Management, and grading system. Additional features are also developed on themes, role management and reporting.

The design of the system is reflected in the Data Flow Diagram (DFD), Entity Relationship Diagram (ERD), and menu interface design before it is developed and implemented.

3.1 User Management

The Users at the new Lentera LMS is integrated into the existing data of students and teachers available at PCU MIS through a syncronization process. The administrator has the access right to do the syncronization before the semester begins. The syncronization takes place in two conditions. Firstly, the offers of each semester courses, including the class schedules have been submitted by each department. Secondly, the process of students's enrollment ends. This syncronization process allows the users to automatically use the Lentera LMS without going through the process of registration to Lentera, as far as they are the members of PCU academic community.

After the syncronization takes place, once the students enroll to a course through PCU MIS, the students will automatically have the access rights to the online course at Lentera using their PCU accounts. Meanwhile, the teachers will also be automatically assigned to the courses they are teaching on the related semesters. Thus, both teachers and students will automatically have the access right to enter into their own courses.

3.2 Course Management

The lists of courses offered each semester at Lentera will be integrated into the database of courses at PCU MIS through the process of syncronization. The administrator has the access right to do the syncronization. The administrator does not need to register the courses manually as far as the courses are documented in the PCU MIS. At Lentera, the courses are grouped under each department at PCU.

3.3 Grading System

The grading system at Lentera is managed by each teacher of the course by assigning grades to the assignments, tests or exams. The Moodle-based Lentera, provides a feature to allow the teachers to give values to the grading variables manually. The calculation of the final grades is based on the method of **weighted mean of grades** as one of the grading aggregate methods of Moodle. Each teacher is allowed to assign weight to each grading variable. An example of calculating the grades using the method of **weighted mean of grades** is as follows:

Variable 1: Test 1 75/100 *weight* 25,

Variable 2: Assignment 60/100 *weight* 15

Variable 3: Midterm Exam 50/100 *weight* 25

Variable 4: Final Term Exam 70/100 *weight* 35

(0.75*25+ 0.6*15 + 0.5*25+0.7*35)/100 = 0.6475 → 64.75/100

The teacher can have an alternative to display or to hide to be viewed by the students. The grades are presented in table form, displaying the detail grades for each grading variable. The management of the grading features by each teacher is utilizing the features provided by Moodle. A new feature is added to allow each teacher to transfer the grades to the university administration system in DBF format.

3.4 Themes

Themes at Lentera follow the themes format available at Moodle. Themes can be customized to represent the use of Lentera LMS at PCU. The selection and management of themes can only be managed by the administrator using his/her access right.

3.5 Role Management

The roles available at this system are grouped into these following categories, each with different access right:

- **Administrator Role** has the highest access right at Lentera. This role has the right to do any modification on all features at Lentera..
- **Course Creator Role** is used to create new courses and arrange the courses at a certain department.
- **Teacher Role** is used to arrange and add contents of a certain coruse. This role is used by each teacher at Lentera.
- **Student Role** is used to participate in the activities of a certain course. This role is used by the students at Lentera.

3.6 Reporting

Reporting implemented in this Lentera LMS is meant for teachers. This feature allows each teacher to view the activity log of each activity designed for each course. Both teachers and students are able to view each detailed activities at Lentera.. This system also provides reports on numbers of access an participation of the students in each course.

4. TESTING AND IMPLEMENTATION

Experimented in 2009, The implementation for this Lentera LMS is currently limited to the courses offered by four departments in the Faculty of Industrial Technology. This consideration is based on the facts that some teachers of these four departments have been used to using Moodle.

The implementation of the LMS involves the installation of several softwares such as WampServer and Moodle. Wampserver is a windows-based application package consisting of some components such as Apache as the web server, , MySQL as the *database server*, and PHP as the programming languange to manipulate the information at the database to display the elements at the web browser.

The interface design of Lentera homepage (http://www.lentera.petra.ac.id) is reflected in Figure 1.

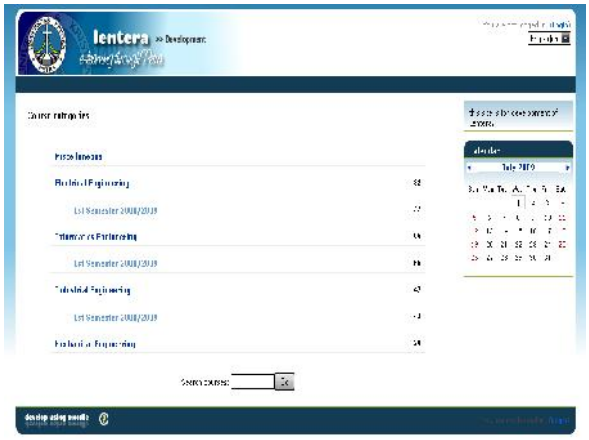


Figure 1. Homepage of Lentera

4.1 Black Box Testing

This testing is only run at the three main features to solve PCU problems in using LMS, the user syncronization, course syncronization and administration and grading system. The additional features are not included.

4.1.1 User Syncronization

The user accounts are managed through the 'users' menu and 'accounts' sub-menu. The addition or modification of the users' data is managed at this sub-menu. The user list can be viewed through the 'browse list of user' menu.

The authentification process is also managed in the '`users`' menu with '`authentication`' as the sub-menu and '`manage authentication`' as the sub-sub-menu. The list of authentification methods available can also be viewed on the above menu.

The syncronization of students' data is managed under the 'MIS synchronization' menu, 'settings' sub-menu, and 'student synchronization' sub-sub-menu. This syncronization is conducted to manage the connection with the students' data available at PCU MIS database. The process of syncronization is reflected at Figure 2 below.

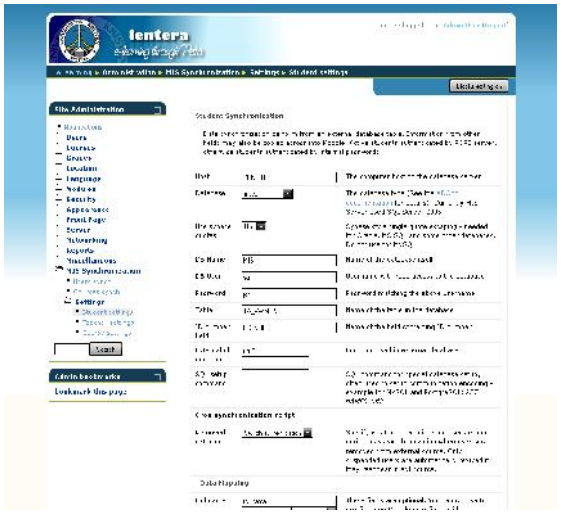


Figure 2. Syncronization on students' data

The syncronization of teachers' data is managed under the 'MIS synchronization' menu, 'settings' sub-menu, and 'teachers synchronization' sub-sub-menu. This syncronization is conducted to manage the connection with the students' data available at PCU MIS database.

The status of syncronization of the students and teachers' data can be viewed under the 'MIS synchronization', menu, and 'users syncronization' menu. Figure 3 shows the connection page.

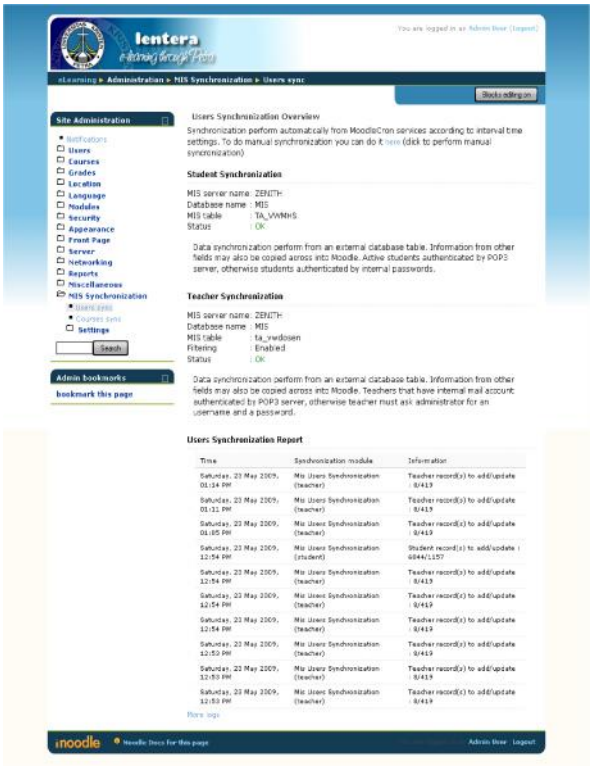


Figure 3. Users Synchronization status

The process of synchronization is automatically run periodically using MoodleCron. Figure 4 shows the number of users at the system before the MoodleCron *service* runs the user synchronization script.

Figure 4. The number of Users before the process of syncronization

Figure 5 shows the number of users at the system after the process of syncronization was run. The process of adding users is run if the users at MIS U.K Petra *database* are not available at the Lentera LMS *database.*

Figure 5. The number of Users after the process of syncronization

Before the process of syncronization, the number of users at the Lentera LMS is 2657. After the syncronization, the number of users are 3531.

4.1.2 Course Syncronization and Administration

The syncronization of courses is managed under the 'courses' menu, 'add/edit courses' sub-menu. Under this menu, the administrator can add or modify the data on courses. The list of courses is catagorized under each department at PCU.

Enrollement is managed under the 'courses' menu and 'enrollments' sub-menu. The list of available enrollment methods that can be used at the Lentera LMS can be viewed. The management of enrollement of the PCU MIS database is conducted to arrange the link between the teachers' class schedules with those of the students at PCU MIS database.

The course participants are managed under 'participants' menu. The list of the course participants can be viewed under this menu. The list of participants can be filtered, based on groups and roles in a course..

The courses are managed under 'settings' menu. Under this menu, the administrator can modify the setting of the courses under the teachers' responsibility. The features include catagorizing courses, assigning names for the courses, description of the courses. Format of teaching-learning process, number of meetings or topics, start and end date of teaching etc.

The course contents can be managed by clicking 'turn editing on' konob at the upper right side of the page or at the left side of the page. A list box will appear and offer alternatives to add resource or activity for each topic or meeting.

4.1.3 Grading System

The grading process is managed under 'grades'menu to display the grader report. The teachers assign grades by activating editing mode by clicking 'turn on editing' knob at the upper right side of the page.

The process of exporting grades is managed under '`grades`' menu. Then a grader report will appear. To export the grades to DBF file at LIS MIS, the teacher can click the '`action`'menu at the list box at the upper right side of the page. During the process of exporting, the teacher can set the value variables for each type of grade and preview the grades to be exported, is presented in Figure 5. Download of the results of the exporting process is conducted using the 'download' knob. A download box will appear to save the files in the web browser.

Figure 5: Preview of Exporting Process

To compare the files with the DBF files of the university administration system, the DBF files resulting from the exporting process is opened using Microsoft Visual FoxPro. The result of the exported grades from Lentera LMS can be seen at Figure 6. While the DBF files of the university administration system can be seen at Figure 7. Both testings give similar results.

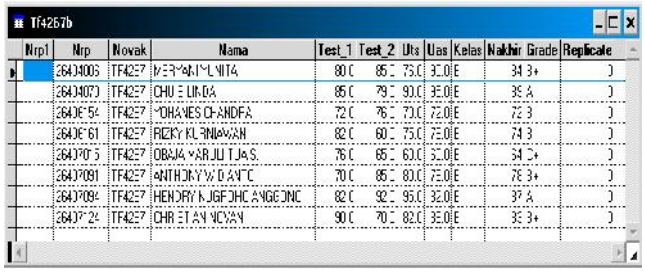


Figure 6: File DBF resulting from the Exporting process from Lentera LMS

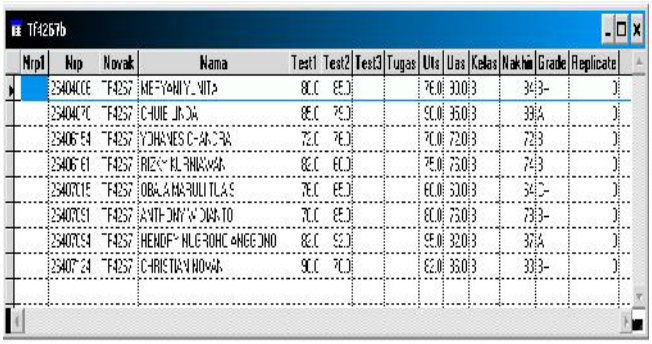


Figure 7. File DBF resulting from Exporting process of university administration system

4.2 USER ACCEPTANCE

This new LMS for PCU was tested against two types of users. Both types of users are from Faculty of Industrial Engineering. The two different sites of Lentera LMS, the system administration site and the course administration site were presented to both types of users. The Questionaires were distributed for responses.

The system administration site was tested against 5 users from different departments who are usually assigned as the LMS administrators or designers. The testings were on the integration aspect and system functional aspect

The responses against the integration system on the system administration site are presented at Table 1. The results indicated that the integration of the students', teachers' and courses' data between those of the Lentera LMS and of PCU MIS ran well and helped the users in using the LMS.

Table 1. Responses against Integration System on System Administration Site

Question \ Respondent	1	2	3	4	5	Average
1	Y	Y	Y	Y	Y	100%
2	Y	Y	Y	Y	Y	100%
Total						100%

The testing for the fungsional aspect of the system *administration site* was using 1-5 scale, which 1 was the highest point (100%), 5 was the lowest point (0 %). The result as indicated at Table 2 shows that the average appraisal was 92 %.

Table 2. Responses against the functional system of the System Administration Site

Question \ Respondent	1	2	3	4	5	Average
3	1	1	1	2	1	96%
4	1	1	2	1	2	92%
5	1	1	2	1	1	96%
6	1	2	2	2	2	84%
Total						92%

The course administration site was tested against 10 teachers from different departments who have been using tthe two PCU LMS. The testing were on system integration, exporting grades, reporting activities and other features of course administration.

The responses against the course administration site are presented at Table 3. The results indicated that the integration of the students', teachers' and courses' data between those of the Lentera LMS and of PCU MIS ran well and helped the users in using the LMS.

Table3. Responses against Integration System on Course Administration Site

Question \ Respondent	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Average
1	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
2	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
Total											100%

The testing for the fungsional aspect of the course *administration site* was using 1-5 scale, which 1 was the highest point (100%), 5

was the lowest point (0 %). The results as indicated at Table 4 shows that the average appraisal was 93,5 % .

Table 4. Responses against the functional system of the Course Administration Site

Respondent / Question	1	2	3	4	5	6	7	8	9	10	Average
3	2	1	1	2	1	1	1	2	1	1	94%
4	1	1	1	1	1	1	1	1	2	1	98%
5	1	1	3	1	1	1	2	2	1	2	90%
6	1	1	2	1	1	2	2	1	1	2	92%
Total											93,5%

5. CONCLUSION

It is concluded that: 1. An accurate setting of PHP is needed to run the process of syncronization, 2. The execution of script for the process of syncronization need 10 – 20 minutes to prevent too early time-out, 3. The process of syncronization needs some adjustments such as field filtering due to different structures of the two databases. 4. Both types of users were satisfied with the results of data integration as well as fungctional aspects of both administration and course administration sites.

6. ACKNOWLEDGMENTS

The authors express many thanks to the related parties such as the PCU Computer Center, PCU MIS Team and Faculty of Industrial Technology academic staff who have contributed to the development of the LMS.

7. REFERENCES

[1] Cole, J., Foster, H. 2008. *Using Moodle* (2nd ed.). Sebastopol: O'Reilly Media, Inc.

[2] Moodle Community 2009. *Moodle docs*. Retrieved 21 January, 2009, from http://docs.moodle

A Web-Based Logistics Information System for Freight Forwarder PT. Rajawali Imantaka Sempurna

Yulia
Petra Christian University
Siwalankerto 121-131
Surabaya
62-31-2983450
yulia@petra.ac.id

Winda Natalia
Petra Christian University
Siwalankerto 121-131
Surabaya

Indro Setiawan
Petra Christian University
Siwalankerto 121-131
Surabaya
62-31-2983950
indro.setiawan@petra.ac.id

ABSTRACT
This paper presents a web-based logistics information system at PT Rajawali Imantaka Sempurna, (PT. RISE), a company engaged in freight forwarding. PT. RISE has two operational offices, in Mataram and Surabaya. With existing legacy systems, they have difficulty to obtain information regarding the shipment of goods and in monitoring the operations of the office. The proposed system is able to handle the freight forwarding process from receipt of order delivery, determining the trucks that will transport the goods, shipping schedule, shipping documents and preparing reports.

Keywords
Logistics, Information System, Freight Forwarding.

1. INTRODUCTION
PT. Rajawali Imantaka Sempurna (RISE) is a company that is engaged in freight forwarder. RISE received a packet delivery order and the wholesale package (not the retail package). The route is from Surabaya-Mataram and Mataram-Surabaya. PT. RISE has two operational offices; the headquarters is located in Mataram and branch office in Surabaya.

Customers will order by phone. Orders will be accepted by the branch manager, then proceed to check the availability of trucks. If the truck is available, the package will be taken at the customer's place. If the truck is not available, the order will be delayed or canceled according to customer demand. The packages will be brought to the office and collected together with other packages. Branch manager will allocate each package to the truck, and administrative officers will calculate the shipping cost of each truck. If the total cost of each truck delivery not worth more than Rp. 8.000.000, then these packages will be allocated again until is worth more than Rp. 8,000,000. Trucks that are ready to depart to the destination accompanied with the Letter of Transport and Cargo Letter.

Payment can be made by the sender or recipient. Once the package arrives at the place, payment officer will provide invoice to the customer. Payment can be made maximum within one month. Payment can be cash or bank transfer. If it is due date, the payment officer will claim the payment by giving the invoice billing. If payment by bank transfer, the customer will provide proof of transfer. Upon payment in full, the payment officer will give payment paid stamp on the invoice payment.

2. LOGISTICS INFORMATION SYSTEM
The investment and adoption of IT is becoming the rule and is essential if forwarders want to survive in the future [1]. The changing competitive conditions have increased pressure on forwarders to compete aggressively and simultaneously on both a low-cost and a differentiated position.

Logistics is the field of study to focus on the design and implementation of the efficient flow and storage of goods from the point of origin to consumption. Information System is the field of study to deal with problems against the design, development, implementation, application of information system. Logistics Information Systems (LIS) is a new discipline that unifies Logistics and Information Systems [2].

3. SYSTEM ANALYSIS
There are some problems with the legacy system are:

a. The lack of information regarding the availability of trucks, so when customers ordered when the truck is not available, the clerk's office cannot provide certainty when the truck will be available.
b. Shipping document still use handwriting, so it could happen obscurity or letter writing is unreadable.
c. Owner has difficulty in monitoring the financial of Surabaya branch office, because the owner stays in Mataram.
d. Difficult to monitor packages in the warehouse have not been sent.
e. There is no monthly performance reporting, such as the number of shipments per month, and the financial statements of income and expenditure.

Based on those problems, the system requirements are as follows:

a. A system that can help to create a simple schedule of availability of trucks.
b. An administrative system that can handle the creating documents problem such as Transport Letter, Cargo Letter, Bill Payment Letter, etc
c. A system that can allow the owner monitors financial expenditure and income at offices in Mataram and Surabaya.
d. A system that can monitor packages that have been sent and has not been sent.

4. SYSTEM DESIGN

4.1 Data Flow Diagram

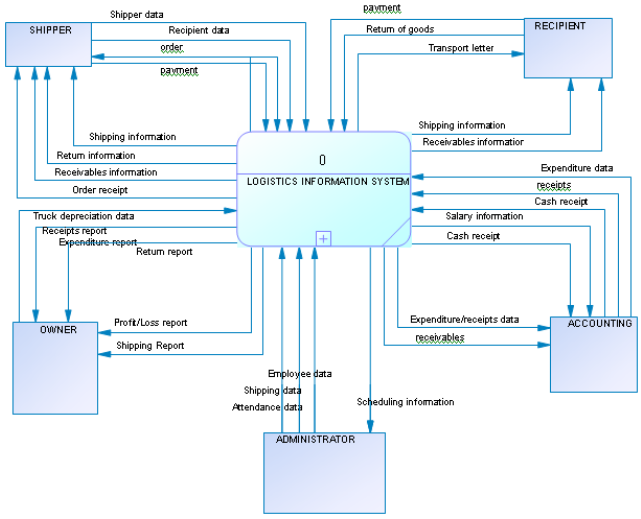


Figure 1. Context Diagram

DFD context diagram (Figure 1) is a portrait of the system in general. The system consists of six entities, namely the sender, receiver, owner, administrator, and finance. The process begins with the order from the sender. The information must be provided is biographical of sender data (if their order is for the first time), the receiver data and package data to be sent. If the order has been collected, the admin will process scheduling shipments.

The process began with determining the delivery date and chooses the shipping destination city. After that, system will search the available truck and driver. If the truck and the driver are not available, then the delivery date has to be changed. If the driver and the truck are available, user must specify the driver and the truck that will be scheduled.

Next, the process of allocating the package into a truck that has been determined will be done automatically by the system using the FCFS (First Come First Served) with priorities, where the system will prioritize orders that come first. If there are orders that come together, the system will choose the order with higher priority. Factors that affect the priority are the number of customer orders ever done. The more often customers use a freight forwarder, the higher the priority.

Package allocation process began with taking the capacity of a truck that has been selected. Then the system will select items that have not been scheduled or sent to the FCFS method. These items will be allocated to the appropriate truck capacity. If the truck is no longer fit, but the cost of shipping freight has not reached the minimum, then the scheduling is invalid and not saved. Scheduling is considered valid if the truck is not greater than the capacity of the truck and the total cost is less than minimum shipping costs. After a valid scheduling, scheduling will be stored in a Scheduling table. Scheduling process can be seen on Figure 2.

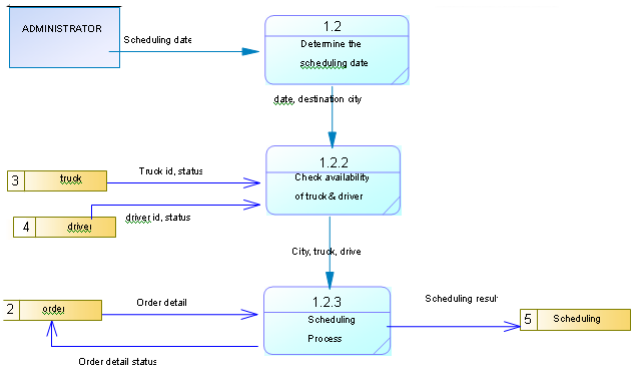


Figure 2. DFD Level 1- Scheduling

The packages that will be shipped are allocated on each truck. Next, it will be continued in the delivery process. Before shipping, admin will process documents such as Transport Letter and Cargo Letter as in Figure 3.

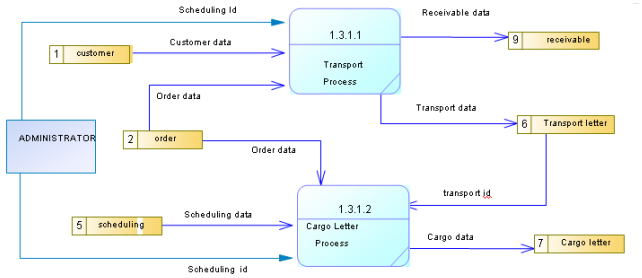


Figure 3. DFD Level 1- Transport & Cargo

After the packets arrive at their destination, bill payments will be given to the party responsible for the shipping costs. The recipient can request return of goods if there are any shipping errors or damage to package, by giving the transportation letter code and items data to be returned.

Every day, the administrator must record employee attendance. The results from the presence of each month will be accumulated with a daily salary of the employee to generate employee salary for one month.

Financial officer will enter data on all expenditures and revenues of each branch office. Revenues came from customer accounts receivable and cash payment of employees, while expenditures came from employee payroll, cash receipt, truck depreciation, and other expenses spent for daily operations. The output of this process is proof of cash receipt, and employee payroll slip. The system will store and generate information in the form of financial statements.

Based on all data being entered and processed by the system, the system will generate output in the form of Transport Letter, Cargo Letter, shipment scheduling, freight costs, availability of truck and driver status, income-expense statements, income statement, delivery report, Report of returns of goods, Report of items that have not been sent, salary slip and salary report.

4.2 Access Right

There are three kinds of permissions, i.e. admin, finance, and the owner. access rights depend on their respective positions in the company and the city where they were. User with access rights Surabaya city, can only see the shipping activity in Surabaya, and the transactions that occurred in Surabaya, as well as for users with access rights Mataram..

Administrator access right can use all menus in the program, unless the financial menu. Administrator can add transaction and view the various data according to the city where they were. The next is the financial access right. User with this access right is allowed to add order and use financial menu accordance to their city.

5. IMPLEMENTATION AND TESTING

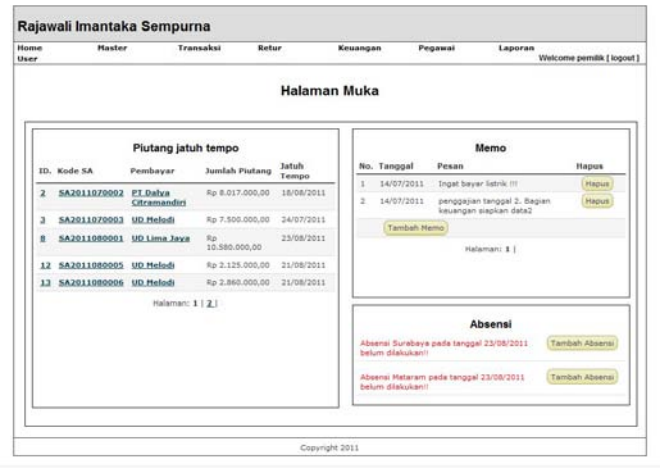


Figure 4. Home Page

Figure 4 is the system's home page, display a warning about the customer receivables that will be due, memo messages, and alerts if there are attendance that have not been done on that date.

Figure 5 is a case study example for a delivery order on August 24, 2011. There is a shipping order from Surabaya. The sender is PT Dalya Citra Mandiri and the recipient is UD Melody, and the payer is UD Melody. Packages to be shipped are 50 packs bowls, 50 packs table, 100 packs Plastic Pails, 100 packs Buckets Louhan. Total cost of shipping is Rp 2.425.000, - and will be due on September 23, 2011. After several orders collected and will be sent, the process continues to scheduling. Scheduling dates can be tailored to the desire shipping date. Destinations can be selected according to delivery destination. After the user specifies the date and select the delivery destination, the system will search for the available driver and truck on the specified date, and in the sender city.

The driver and the truck was registered to another scheduling cannot be selected again. In addition, the driver and the truck was registered in the sending process until the date of scheduling, it cannot be selected too. If there are no trucks and drivers are available, the scheduling cannot be done. Scheduling can be saved and considered valid when the weight of the goods delivered no more than the capacity of a truck with delivery charge no less than the minimum delivery charges.

Scheduling for the delivery of goods made on August 25, 2011. Delivery of goods to the city of Mataram by using a truck with police number L 1234 AB, and driven by a driver named Iwan. Here are the results of scheduling, can be seen on Figure 6.

Shipping recording process began by selecting the desired schedule, and then will display the details of scheduling that has been selected. This feature can only be performed by a user with admin or owner access rights Users with admin access right can only make delivery of package from the city it is located, while the owner access right may make delivery of goods from any city.

On August 24, 2011 will be made a package delivery for scheduling number 1 to Mataram. Results of Transport letter (*Surat Angkutan*) can be seen on Figure 5 and Shipping (*Pengiriman*) Form can be seen on Figure 6.

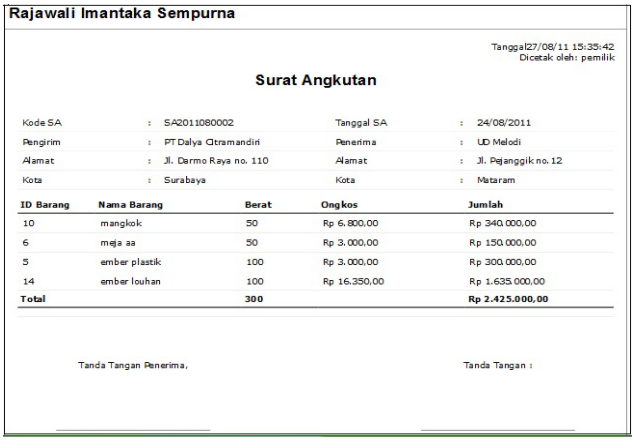


Figure 5. Transport Letter *(Surat Angkutan)*

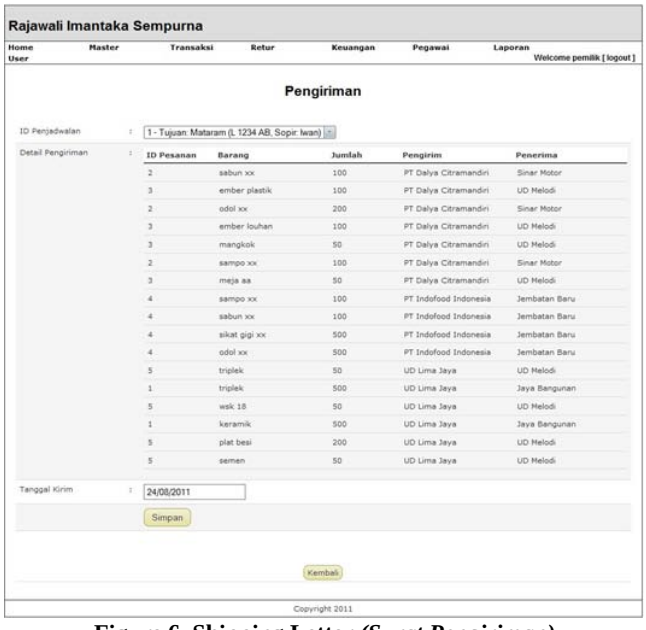


Figure 6. Shipping Letter *(Surat Pengiriman)*

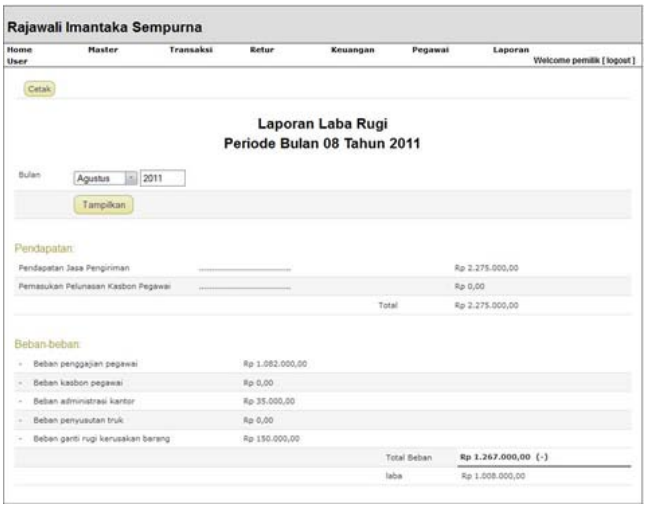


Figure 7. Income Statement *(Laporan Laba Rugi)*

Income Statement on Figure 7 is a report that displays the company's overall revenue and expenditure, and company's profit or loss statement for a particular month. This report can only be accessed by users with owner's access rights.

6. CONCLUSION

At the end of the design and implementation of web-based information system of PT RISE, it can be concluded that:

- Scheduling system is less efficient because of the scheduling system can only calculate the charge based on weight and not by volume of the package.
- With this online application, the owner can monitor the shipment activities.
- Based on the results of the questionnaire, this application gets a percentage of the overall average of 82.5% so it can be said that this application is sufficient to meet the needs of the company.

7. REFERENCES

[1] Hardaker, G., Trick, R.R. and Sabki, A.A. 1994, The Use of IT in Freight Forwarding in the UK, International Journal of Logistics Information Management, **7.4**, 19-22.

[2] http://www.slideshare.net/togar/logistics-information-system, 8 March 2012

Decision Support System for Supplier Selection by Using Analytic Network Process (ANP) Method for the Procurement Department

Alexander Setiawan
Petra Christian University
Siwalankerto 121 – 131 Surabaya
60236 Indonesia
Telp. (031) – 2983455
Fax (031) – 8417658
alexander@peter.petra.ac.id

Leo Willyanto Santoso
Petra Christian University
Siwalankerto 121 – 131
Surabaya 60236 Indonesia
Telp. (031) – 2983455
Fax (031) – 8417658
leow@peter.petra.ac.id

Margaretha Juan
Petra Christian University
Siwalankerto 121 – 131 Surabaya
60236 Indonesia
Telp. (031) – 2983455
Fax (031) – 8417658
m26405011@john.petra.ac.id

ABSTRACT

In this life, every each person has to make decision in every moment. Making Decision is not an easy task to do. There are always some consequences in making decision. Most of us don't know what the best decision is until we make it. In business life, company as an organization also has to make decisions. One of the most important is making decision for employee recruitment. The company's performance depends on the human resource's performances. Development team has some problems in this case. There are a lot of subjectivities when the team wants to make decision in supplier selection.

Based on that fact, the company realizes that they need a support computer application in making decision especially for the employee recruitment. Development team needs supporting application using Analytic Network Process. The application will be developed to provide suggestions to the manager.

Keywords:

Decision Support System, Analytical Network Process, Supplier, Criteria.

1. INTRODUCTION

The procurement department was the unit that was under the Unit Perencanaan Fisik Kampus Unit (UPFK) "PT X" that was assigned in handling the tender for the thing or the service. The procurement part acted as the organizer of the tender in satisfying the requirement for "PT X" in large quantities. In holding the tender, the procurement department will carry out the selection of the candidate supplier to be invited. All the candidates supplier that was invited to have the appearance of his company that was good and in accordance with the standard. The procurement department was assigned in inviting the candidate supplier, held the tender and broke supplier that won. The candidate's selection supplier was carried out was based on bargaining compatibility towards the available requirement, the experience from the past, the available recommendation, the experience from the candidate's client supplier previous, as well as the price that was offered.

In the process of the selection of the candidate supplier, the decision that was taken often was influenced subjectivity from the decision-makers. This was also caused by the assessment of the candidate supplier until now was not done objectively because of the nonexistence of the standard in the assessment. Subjectivity could happen because of not having the standard method that was systematic to consider the appropriateness of the candidate supplier. One of the decision making methods that could be used in the process of the selection of the candidate supplier was the Analytic Network Process method (ANP). ANP was chosen as the method in this research because of could overcome relations between sub the criterion.

Especially the formulation of the problem in this research was:

- How made the database plan that could keep the thing data, the candidate supplier, the criterion, the sub-criterion, and the value of the relations that happen.
- How made an application that could help take the decision in helping determined the candidate supplier for each thing.

The aim of this research was the based application web to support the process of decision making in chose supplier each thing by making use of the Analytic Network System method (ANP), so as to be received by the alternative to the decision of the election supplier the thing objectively as well as in accordance with the criterion that was needed.

2. FOUNDATION OF THEORY

2.1 Analytical Network Process (ANP)

The Analytic Process method (ANP) was the development of the Analytical Hierarchy Process method (AHP). The ANP method of could improve the AHP weakness took the form of the capacity the connection between the criterion or the alternative (Saaty, 2004). The connection in the ANP method of having 2 kinds that is the connection in one set of the element (inner dependence) and the connection between the different element (outer dependence). The existence of this connection caused the ANP method more complex compared to the AHP method.

The step in using ANP was began decisively the criterion and all of his weights afterwards ANP began counted after all the criteria had the weight. Super the matrix weighted and cluster the matrix was received from input the criterion weight whereas weighted was received with multiplied weighted with cluster and limiting the matrix was received with weighted the matrix to the value each one of his lines became same.

Table 1. The comparison matrix

A	B_1	B_2	B_3	---	B_n
B_1	b_{11}	b_{12}	b_{13}	---	b_{1n}
B_2	b_{21}	b_{22}	b_{23}	---	b_{2n}
B_3	b_{31}	b_{32}	b_{33}	---	b_{3n}
---	---	---	---	---	---
B_n	b_{n1}	b_{n2}	b_{n3}	---	b_{nn}

Thought bij was the Bi value of the element comparison against Bj that stated relations:

- As far the level of the Bi interests when compared with Bj, or
- It was as big that the Bi contribution towards the A criterion compared with Bj, or
- As far the Bi domination compared with Bj, or
- How many characteristics of the A criterion were received to Bi compared with Bj.

When being known by the value bij then theoretically thought bji = 1/bij, whereas bij in the situation i = j was absolute 1.

The numeric value that was used for the comparison above was received from the scale of the comparison that was made by Saaty (2004). The table 2 determined the scale of the comparison between the elements in the process of decision making.

Table 2. The assessment of pair wise comparisons

Priority	Definition	Note
1	As Important	The two elements had the same influence
2	Weak or Slight	
3	Slightly more Important	Eperience and judgement slightly favor one activity over another
4	Moderate Plus	
5	More Important	Eperience and judgement strongly favor one activity over another
6	Strong Plus	
7	Very Important	An activity is favored very strongly over another
8	Very, Very Strong	
9	Absolutely Important	One element was proven to be absolute more was liked compared with his couple

2.2 The Calculation Process of ANP

Weighted with ANP needed the model that mutual connections between the criterion and the sub-criterion that were owned by him. There were 2 controls that must pay attention to inside the system that will be known by his weight. The first control was the control of the hierarchy that showed the criterion connection and sub his criterion. In this control did not need the structure of the hierarchy like in the method of AHP. Control other was the control of the connection that showed the existence of mutual connections between the criterion or cluster (Saaty, 2004). The influence from one set of the element in some cluster in the element that was other in a system could trough the priority vector on a scale the ratio that was taken from the teamed comparison. The network in this method of having the complexity that was high compared with the other kind, because of the existence of the phenomenon feedback from cluster one to cluster other, in fact with cluster him himself. The criterion of the official's candidate was stated as cluster whereas the element and sub his element was the objective strategy. To the Figure 1., showed the network model with feedback and dependence cluster one with cluster other.

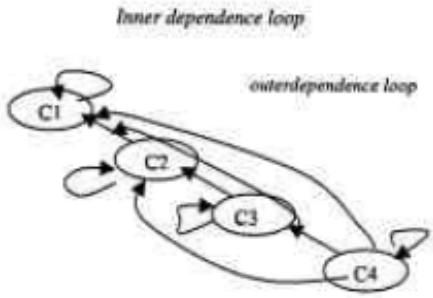


Figure 1. Feedback Model and Dependence Cluster (Saaty, 2004)

After the model was made, then was carried out from results of the data pair wise comparison by using the super-matrix table. To the Figure 2 were shown by the format of the foundation of the super-matrix table.

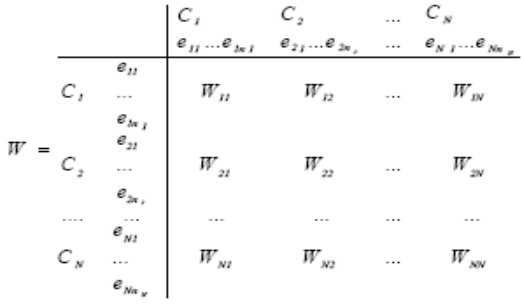


Figure 2. The format of the Foundation Table the Super-Matrix (Saaty, 2004)

Where the bloc i and j from this matrix was:

$$W = \begin{bmatrix} W_{i1}^{(j_1)} & W_{i1}^{(j_2)} & \cdots & W_{i1}^{(j_n)} \\ W_{i2}^{(j_1)} & W_{i2}^{(j_2)} & \cdots & W_{i2}^{(j_n)} \\ \cdots & \cdots & \cdots & \cdots \\ W_{in_i}^{(j_1)} & W_{in_i}^{(j_2)} & \cdots & W_{in_i}^{(j_n)} \end{bmatrix}$$

Figure 3. Matrix Blok *i* dan *j* (Saaty, 2004)

After did the process above, afterwards will be carried out by the process weighted to every time cluster that was determined was based on the criterion of the official's candidate. The calculation algorithm weighted like in the Equality 2.1. that was carried out was begun from the data with the form pair wise comparison to was produced by the weight of each indicator of his achievement. The criterion was made be based on the requirement and the aim of the election.

To show results of the end of the calculation of the comparison then the super-matrix will be promoted continually to the figure of each column in one equal line.

$$\lim_{M \to \infty} \frac{1}{M} \sum_{k=1}^{M} \frac{\sum_{j=1}^{n} a_{ij}^{(k)}}{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{(k)}} \qquad \text{..... (2.1)}$$

3. DISCUSSION AND ANALYSIS

3.1 The Analysis and Design System

Steps that were carried out in this research were as follows :

- A Literature Review
  This step was carried out by studying theories that were used in the production of this thesis.
- Analysis and Design System
  In this stage was carried out by the identification and the evaluation of the problem that happened, as well as looked for the solution from this problem. After the analysis stage was finished was carried out, was made by design planning of the system on the whole.
- Create the Software
  Implement the design of the system that has been made in software. The production of this software covered the whole form input and output, the production of the interface website that was good, as well as the connection with the database.
- Test the Software
  The testing was carried out with put the data inside form. This data was afterwards processed by the system to produce output that in accordance with the requirement. Compared results of the calculation of the manual with results of the calculation of the program.
- Decision Making
  This step was carried out by means of compared whether the production of this software could answer the requirement for the selection of the election supplier the thing .

Was based on observation that was carried out towards the selection system of the candidate supplier the procurement department of "PT X", the available problem that is for the selection supplier did not yet be fully in accordance with the criterion that was determined, the assessment was more often carried out was based on the experience so as the assessment that was made be subjective. Moreover was not yet available data collection that was good for recorded the data of the candidate's criterion supplier because the data that was owned at this time was still taking the form of the document was not written

From the problem that emerged, then several matters along with this was needed by the procurement department in the production of the supporting system of decision making, in part:

- Was needed a system with the database concerning the candidate's data supplier. So as if the tender company could make use of the data that was owned beforehand.
- Explained classification of his database
- Was needed a system that could support decision making objectively in the selection of the candidate supplier by means of putting forward the value of the comparison between the candidate as well as compatibility towards the criterion that was needed.

Context Diagram to the Figure 4 will give the explanation about the data current globally in the system. The available data whole will head to one big process that is the process system mining. Entity Relationship Diagram (Conceptual Data Model) to the Figure 5.

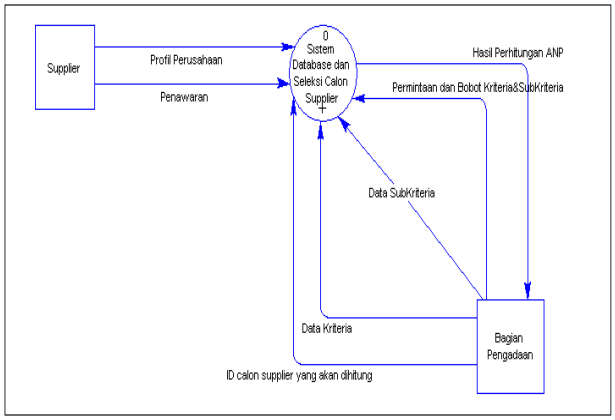


Figure 4. Context Diagram

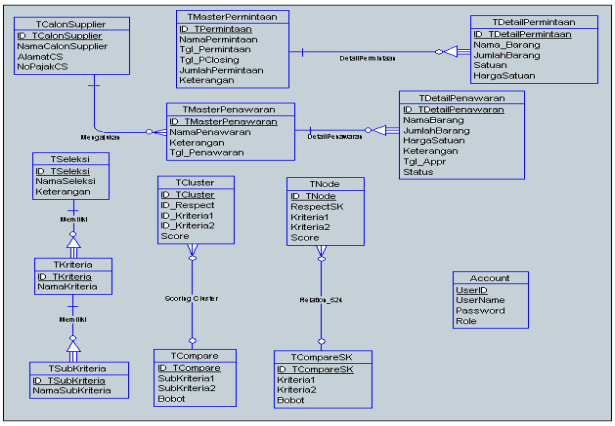


Figure 5. ERD Conceptual Data Model

3.2 The Implementation System

The testing of this system was begun from the page of early that will be used to enter the page of the menu. The user who was registered must put the user id and password to this page to be able to access the following page. The page of the system and the menu of the beginning main could be seen in the Figure 6.

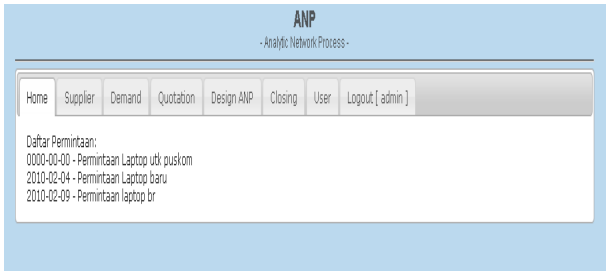


Figure 6. Main Menu

To this page the user could put the list of the request that will be tendered by filling up each available column completely. The request could be seen in the Figure 7. The page of the candidate supplier bargaining can be seen at Figure 8.

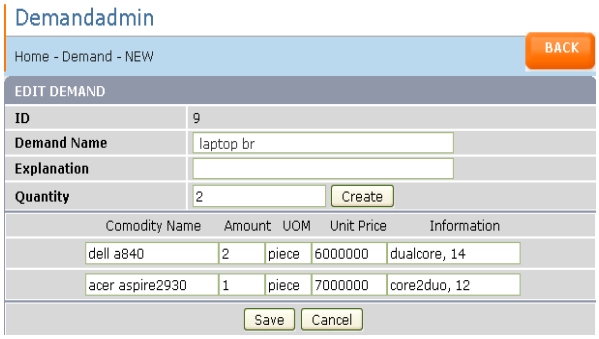


Figure 7. Form Request Input Page

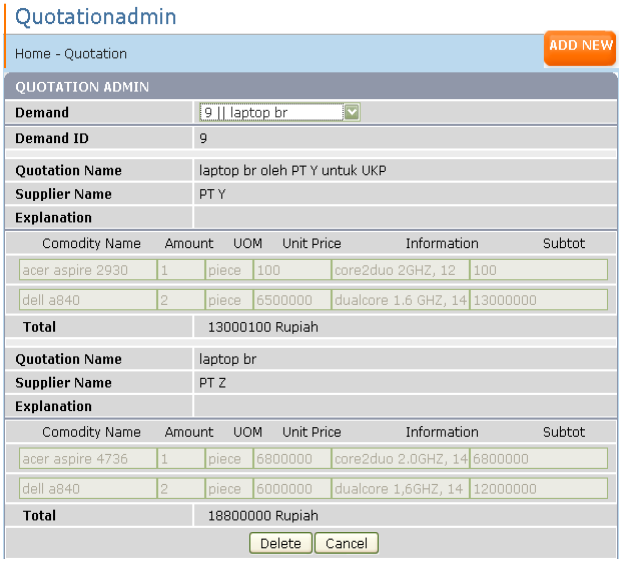


Figure 8. The page of the Candidate's Supplier Bargaining

This page put forward the appearance as well as bargaining that had been done by supplier that was registered. The user could put the criterion and the sub-criterion of the selection election in accordance with the requirement. The criterion and the sub-criterion could be seen in the Figure 9 and the Figure 10.

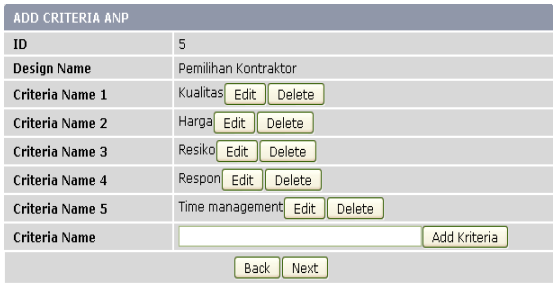


Figure 9. Criteria input Page

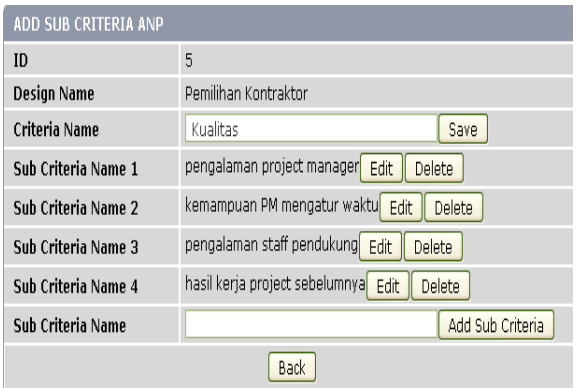


Figure 10. Sub criteria input page

The user could do connected between the sub-criterion so as to mutual dependences between the sub-criterion happen as well as did weighted towards the criterion and the sub-criterion that had relations of mutual dependences. The page of the connection could be seen in the Figure 11. and the criteria value page could be seen in the Figure 12. and the alternative criteria page could be seen in the Figure 13.

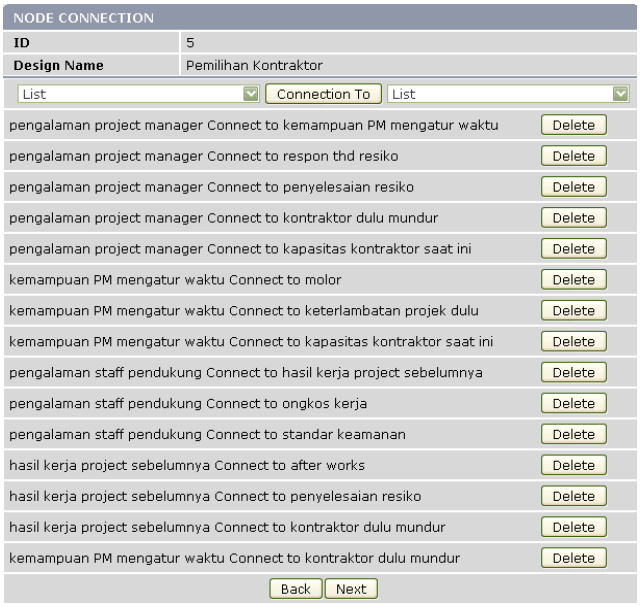


Figure 11. Connection between sub criteria page

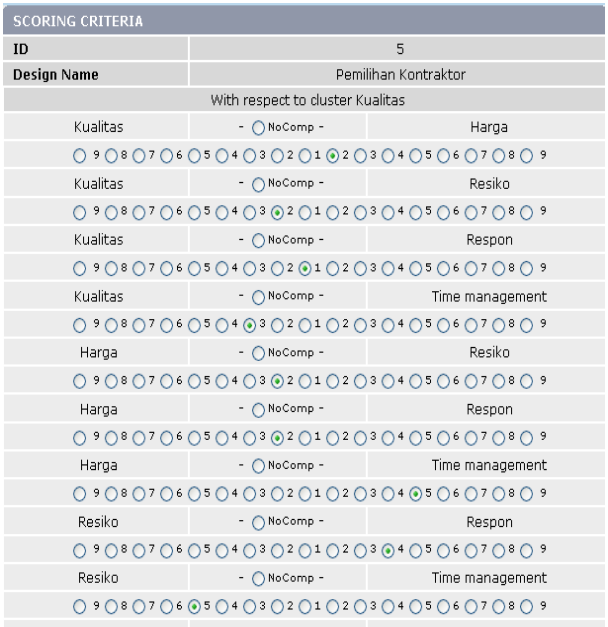


Figure 12. Criteria valuing page

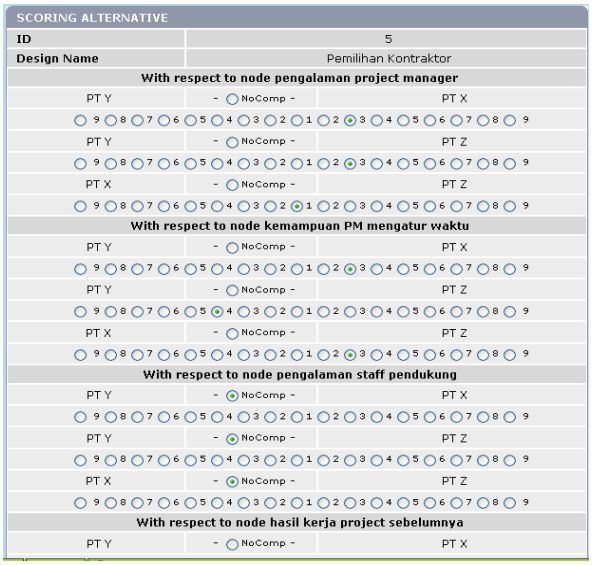


Figure 13. Alternative valuing Page

Results of the calculation weighted the super-matrix manually with the system showed same results so as the calculation of the system could be stated valid but the calculation In A Super Manner Decision 1,6,0 had several differences that could be caused by the difference each result of the calculation that happened. The page of weighted super matrix could be seen in the Figure 14.

unwighted matrix:

	PT Y	PT X	PT Z	pengalaman project manager	kemampuan PM mengatur waktu	pengalaman staff pendukung	hasil kerja project sebelumnya	harga material	ongkos kerja	standar keamanan
PT Y	0	0	0	0.1428	0.3541	0.3333	0.5889	0.6	0.3699	0.3333
PT X	0	0	0	0.4286	0.3331	0.3333	0.1593	0.2	0.2979	0.3333
PT Z	0	0	0	0.4286	0.3129	0.3333	0.2518	0.2	0.3323	0.3333
pengalaman project manager	0.2333	0.1356	0.1357	0	0	0	0	0	0	0
kemampuan PM mengatur waktu	0.2156	0.3333	0.3753	1	0	0	0	0	0	0
pengalaman staff pendukung	0.2923	0.2317	0.258	0	0	0	0	0	0	0
hasil kerja project sebelumnya	0.2588	0.2994	0.231	0	0	1	0	0	0	0
harga material	0.5	0.25	0.5	0	0	0	0	0	0	0
ongkos kerja	0.5	0.75	0.5	0	0	1	0	0	0	0
standar keamanan	0.1977	0.2934	0.2507	0	0	1	0	0	0	0
after works	0.232	0.1283	0.183	0	0	0	1	0	0	0
molor	0.2117	0.2621	0.1637	0	1	0	0	0	0	0
pembengkakan biaya	0.1703	0.1671	0.2129	0	0	0	0	0	0	0
perubahaan	0.1882	0.149	0.1898	0	0	0	0	0	0	0
respon thd resiko	0.5	0.3333	0.875	0.75	0	0	0	0	0	0
penyelesaian resiko	0.5	0.6667	0.125	0.25	0	0	1	0	0	0
keterlambatan projek dulu	0.1638	0.2	0.6901	0	0.3338	0	0	0	0	0
kontraktor dulu mundur	0.2972	0.4	0.2447	0.2	0.1416	0	1	0	0	0
kapasitas kontraktor saat ini	0.539	0.4	0.0653	0.8	0.5247	0	0	0	0	0

Figure 14. Weighted Super matrix

Cluster matrix will be used to give weighted in weighted the super-matrix so that to weighted super matrix. Cluster matrix was compiled from eigen vector the influence comparison cluster could be seen in the Figure 15.

Cluster matrix:

	_Alternative	Kualitas	Harga	Resiko	Respon	Time management
_Alternative	0	0.1667	1	1	1	1
Kualitas	0.1499	0.1587	0	0	0	0
Harga	0.2071	0.0433	0	0	0	0
Resiko	0.2408	0.2335	0	0	0	0
Respon	0.1808	0.1992	0	0	0	0
Time management	0.2215	0.1986	0	0	0	0

Figure 15. Calculate Eigen Vector Cluster matrix

Weighted super matrix will be counted to weighted the super-matrix by means of carrying out multiplication between the contents weighted the super-matrix and cluster matrix. To the Figure 16 and the Figure 17 will be shown limiting Super- Matrix that was produced by Super Decision 1,6,0.

Super Decisions Main Window: kontraktor.mod: ratings: Limit Matrix

	PT X	PT YY	PT Z	harga m~	ongkos ~	hasil k~	kemampu~	pengala~	pengala~	after w~
PT X	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918
PT YY	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373
PT Z	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827
harga m~	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470
ongkos ~	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068
hasil k~	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258
kemampu~	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577
pengala~	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223
pengala~	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707
after w~	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160
molor	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570
pembeng~	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332
perubah~	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300
standar~	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768
penyele~	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312
respon ~	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746
kapasit~	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493
Keterla~	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547
kontrak~	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354

Figure 16. Limiting Super Matrix Super Decision 1

molor	pembeng~	perubah~	standar~	penyele~	respon ~	kapasit~	Keterla~	kontrak~
0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918	0.14918
0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373	0.15373
0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827	0.13827
0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470	0.02470
0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068	0.06068
0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258	0.03258
0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577	0.03577
0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223	0.02223
0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707	0.02707
0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160	0.02160
0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570	0.01570
0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332	0.01332
0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300	0.01300
0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768	0.02768
0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312	0.09312
0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746	0.03746
0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493	0.04493
0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547	0.04547
0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354	0.04354

Figure 17. Limiting Super matrix Super Decision 2

4. CONCLUSION

The purpose of this research of the supporting system of the based decision web in the procurement department of "PT X" in the process of the selection supplier by using the Analytical Network Process method, could be concluded that:

- Results of the ANP calculation that was carried out in this application in accordance with results of the ANP calculation theory.
- The application system that was developed could help the administrative the procurement department of "PT X" in did data collection in a manner so as to support the balanced and objective assessment.
- This research could help the candidate's data collection supplier the procurement department of "PT X" because of being supported with the database arrangement that structured.

5. REFERENCES

[1] Ihsan, Y. 2007. *Personel selection using analytic network process.* Istanbul : İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl.

[2] Liauwnoto, S. 2009. Model pemilihan kontraktor pemenang tender di sektor swasta menggunakan *analytical hierarchy process* (AHP), Tesis Program Pascasarana Magister Teknik Sipil Universitas Kristen Petra.

[3] McLeod, R. 2008. *Management Information Systems*. 10th Edition, New Jersey : Prentice Hall, Inc.

[4] Saaty, T.L. 2004. Fundamentals of the analytic network process dependence and feedback in decision-making with a single network. Pittsburgh : RWS Publications.

[5] Saaty, T.L. 2008. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. Pittsburgh : RWS Publications.

[6] Sebnemburnaz & Y. Ilker Topcu. 2006. A multiple-criteria decision-making approach for the evaluation of retail location, journal of multi-criteria decision analysis, Wiley InterScience

[7] Vanany, I. 2003. Aplikasi analytic network process (ANP) pada perancangan sistem pengukuran kinerja (Studi Kasus pada PT. X), Jurnal Teknik Industri, vol 5 (No. 1-10) Institut Teknologi Sepuluh Nopember. Retrived November 11, 2008, from http://puslit.petra.ac.id/journals/industrial

Web Page Similarity Searching Based on Web Content

Gregorius Satia Budhi
Informatics Department
Petra Chistian University
Siwalankerto 121-131
Surabaya 60236, Indonesia
(62-31) 2983455
greg@petra.ac.id

Justinus Andjarwirawan
Informatics Department
Petra Christian University
Siwalankerto 121-131
Surabaya 60236, Indonesia
(62-31) 2983455
justin@petra.ac.id

Rubia Sari Setiadi
Informatics Department
Petra Christian University
Siwalankerto 121-131
Surabaya 60236, Indonesia
(62-31) 2983455

ABSTRACT
Application that discussed in this paper is able to perform the process of finding web pages that have similar content to the url of the desired web page. Also developed an automated process for crawling web pages. This crawling process will continue since the process is activated. The search process begins by entering a url and web page url is obtained from the extract to get the key words that represent the web page. The keywords will be processed into a basic form using the Porter Stemmer algorithm. TF-IDF method used to obtain the importance of a keyword. Furthermore Jaccard Coefficient formula used to find similarity between web pages. Applications are limited to Web Page in English. Based on test results concluded that this application has worked well and can be utilized.

Keywords
Web Page Similarity, Crawler, TF-IDF, Porter Stemmer, Jaccard Coefficient, Keyword Extraction

1. INTRODUCTION
It has been widely available search engines to find data. No doubt, internet is the largest current source of data. Realizing these facts that the available data increases rapidly, then we can conclude tremendous potential to find data via the web. The web page has a variety of content, but sometimes there are web pages that discuss the same event.

The application will do a search on the web page that has similarities to other web pages. The similarity search based on the contents of the web pages.

2. SUPPORTING THEORY
The only way to collect URLs is to scan collected pages for hyperlinks to other pages that have not been collected yet. This is the basic principle of crawlers. They start from a given set of URLs, progressively fetch and scan them for new URLs (out-links), and then fetch these pages in turn, in an endless cycle. New URLs found thus represent potentially pending work for the crawler [3].

2.1 Automatic Keyphrase Extraction
We define automatic keyphrase extraction as the automatic selection of important, topical phrases from within the body of a document.

Many journals ask their authors to provide a list of keywords for their articles. We call these keyphrases, rather than keywords, because they are often phrases of two or more words, rather than single words. We define a keyphrase list as a short list of phrases (typically five to fifteen noun phrases) that capture the main topics discussed in a given document [7].

2.2 Stopword and Stemming
Stopword [3]: Most natural languages have so-called function words and connectives such as articles and prepositions that appear in a large number of documents and are typically of little use in pinpointing documents that satisfy a searcher's information need.

Stemming [3]: Device to help match a querry term with a morphological variant in the corpus. Common stemming methods use a combination of morphological analysis and dictionary lookup. Stemming ca increase the number of documents in the response, but may at times include irrelevant documents.

2.3 Porter Stemmer Algorithm
Here we present the Porter Stemmer algorithm (Suffix Stripping Algorithm) that we use in the application:

To present the suffix stripping algorithm in its entirety we will need a few definitions [5].

A \consonant\ in a word is a letter other than A, E, I, O or U, and other than Y preceded by a consonant. (The fact that the term `consonant' is defined to some extent in terms of itself does not make it ambiguous.) So in TOY the consonants are T and Y, and in SYZYGY they are S, Z and G. If a letter is not a consonant it is a \vowel\.

A consonant will be denoted by c, a vowel by v. A list ccc... of length greater than 0 will be denoted by C, and a list vvv... of length greater than 0 will be denoted by V. Any word, or part of a word, therefore has one of the four forms:

CVCV ... C

CVCV ... V

VCVC ... C

VCVC ... V

These may all be represented by the single form

[C]VCVC ... [V]

where the square brackets denote arbitrary presence of their contents.

Using (VC){m} to denote VC repeated m times, this may again be written as

[C](VC){m}[V].

m will be called the \measure\ of any word or word part when represented in this form. The case m = 0 covers the null word.

The \rules\ for removing a suffix will be given in the form

(condition) S1 -> S2

This means that if a word ends with the suffix S1, and the stem before S1 satisfies the given condition, S1 is replaced by S2. The condition is usually given in terms of m, e.g.

(m > 1) EMENT ->

Here S1 is `EMENT' and S2 is null. This would map REPLACEMENT to REPLAC, since REPLAC is a word part for which m = 2. The `condition' part may also contain the following:

*S - the stem ends with S (and similarly for the other letters).

v - the stem contains a vowel.

*d - the stem ends with a double consonant (e.g. -TT, -SS).

*o - the stem ends cvc, where the second c is not W, X or Y.

And the condition part may also contain expressions with \and\, \or\ and \not\, so that

(m>1 and (*S or *T))

tests for a stem with m>1 ending in S or T, while

(*d and not (*L or *S or *Z))

tests for a stem ending with a double consonant other than L, S or Z. Elaborate conditions like this are required only rarely.

In a set of rules written beneath each other, only one is obeyed, and this will be the one with the longest matching S1 for the given word. For example, with

SSES -> SS

IES -> I

SS -> SS

S ->

(here the conditions are all null) CARESSES maps to CARESS since SSES is the longest match for S1. Equally CARESS maps to CARESS (S1=`SS') and CARES to CARE (S1=`S').

2.4 Term Frequency - Inverse Document Frequency

Tf-idf method is a way to give weight to the relationship of a word (term) of the document. This method combines the two concepts for calculating weights: frequency of occurrence of a word within a particular document and the inverse frequency of documents containing the word. Frequency of occurrence of the word in the document are given showing how important word in the document. Frequency of documents containing those words show how common the word [4].

The general formula for tf-idf:

$$w_{ij} = tf \times idf$$

$$w_{ij} = tf_{ij} \times \log\frac{N}{n} \quad (1)$$

Description:

w_{ij} = weight of the word / term t_j in the document

tf_{ij} = number of occurrences of the word / term t_j in the d_i

N = number of all documents in the database

n = number of documents containing the word / term t_j

Based on the above formula, regardless of the value of tf_{ij}, if N = n then we will get the result 0 (zero) for the calculation of the IDF. It can be added to the value 1 in the idf, so the calculation of the weight to be as follows:

$$w_{ij} = tf_{ij} \times (\log(N/n) + 1) \quad (2)$$

In this paper the calculation of the tf will be replaced by the calculation method that we proposed in another paper [2].

2.5 Weight of Word / Term

Weight 1 (W1) is the frequency of words in a article. The number of same words in the article is calculated. The result will be divided by the total words in the article, by also considering the frequency of the words in the article [2].

Weight 2 (W2) is a value that is determined by the position of a first sentence that is used the word in a paragraph. In general, every paragraph in a good writing of an article usually only provides one main idea [2]. Because this application is used to process documents in English, then we use the formulation from Jonas and Araki to calculate W2, namely: W2 = Early(j) = 2 if j < 10 first sentence in a paragraph and 1 otherwise [6].

The calculation of the tf to be as follows:

$$tf_{ij} = W1_{ij} \times W2_{ij} \quad (3)$$

2.6 Jaccard Coefficient

The percentage of relevance covered by two sets is known as the Jaccard coefficient and is given by

$$sim(q, d_j) = J(A, B) = \frac{|A \cap B|}{|A \cup B|} \cong \frac{\sum_{k=1}^{n} w_{kq} w_{kj}}{\sum_{k=1}^{n} w_{kq}^2 + \sum_{k=1}^{n} w_{kj}^2 - \sum_{k=1}^{n} w_{kq} w_{kj}} \quad (4)$$

This measure is fairly intuitive and often one of the more widely used measures when comparing IR systems. In a set theoretic

sense, the Jaccard measure signifies the degree of relevance covered by the union of two sets [1].

3. APPLICATION DESIGN

The design of the the application can be seen in the flowchart in Figure 1 to Figure 5.

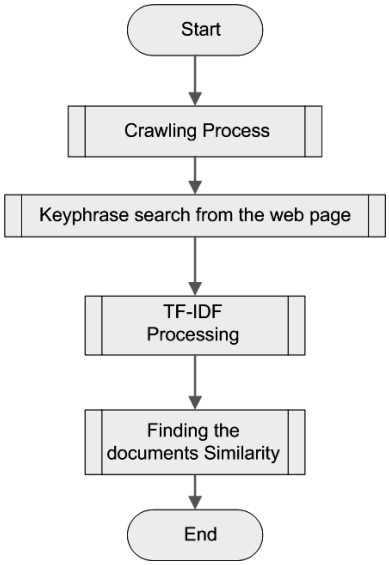


Figure 1. Application Design Flowchart

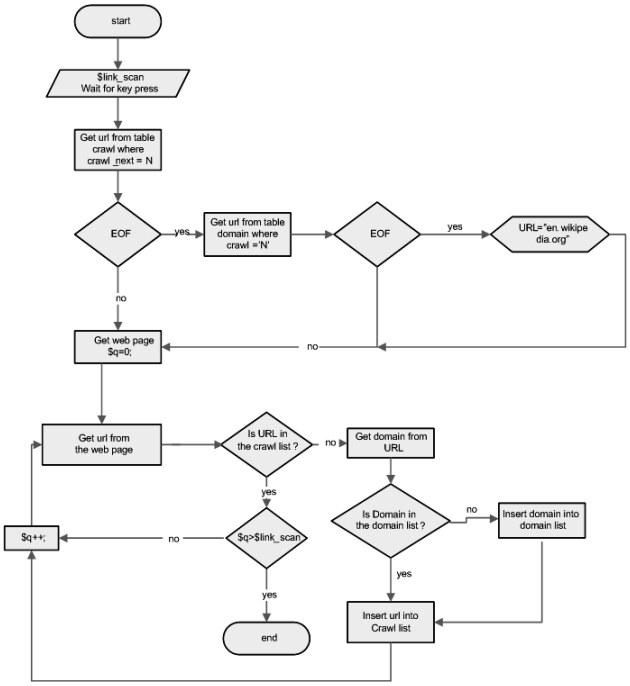


Figure 2. Crawling Process Flowchart

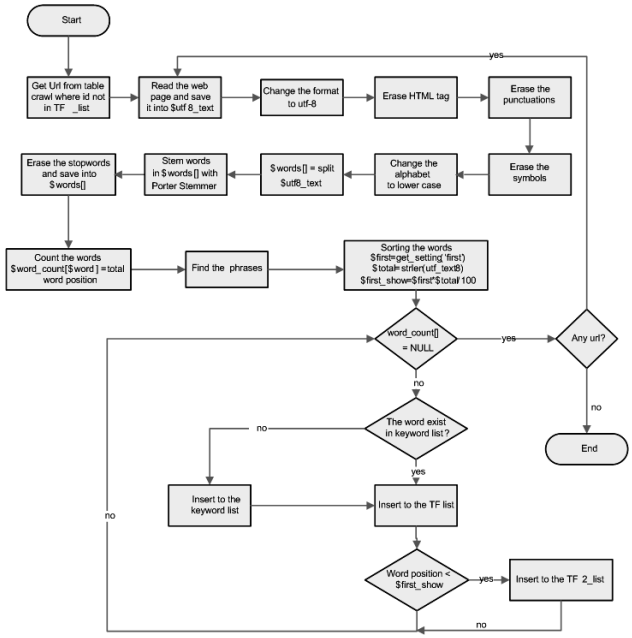


Figure 3. Keyphrases Search Flowchart

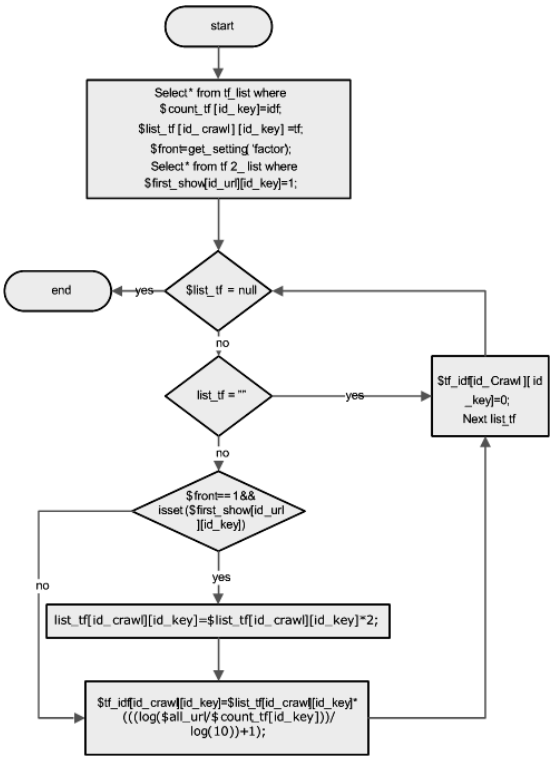


Figure 4. TF-IDF Processing Flowchart

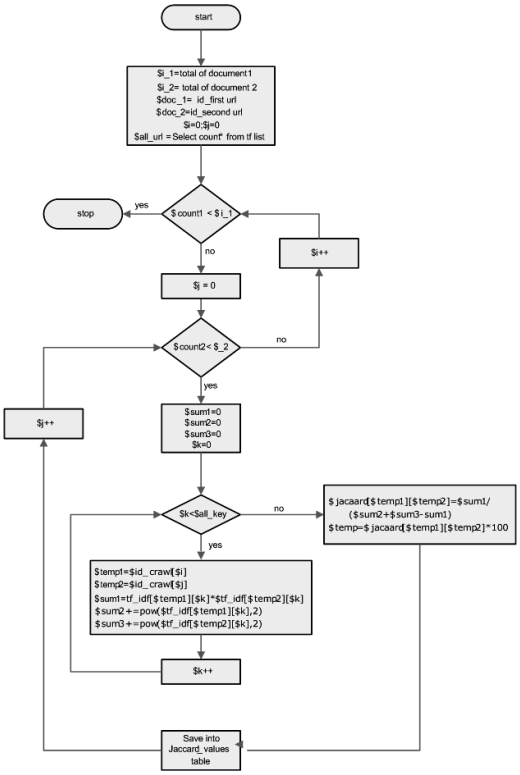


Figure 5. Document Similarity Finding Flowchart

4. APPLICATION INTERFACES

The interfaces of the application are divided into two parts, namely:

a. Pages for the User:
On the main page (Figure 6) the user can enter a web page to search its similarity to another website. After pressing 'enter' or press the 'search' button then the application will find and display similar web pages. The Display results can be seen in Figure 7.

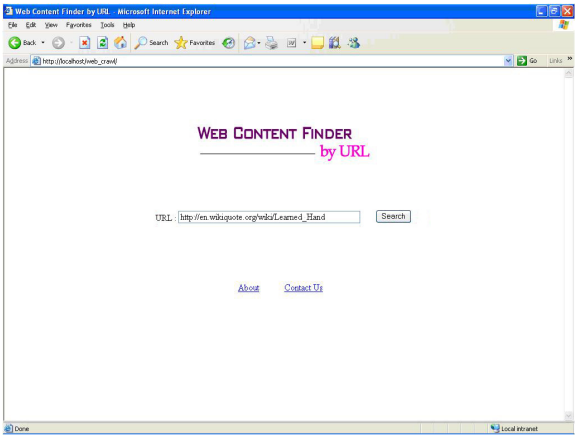


Figure 6. User Main Page Interface

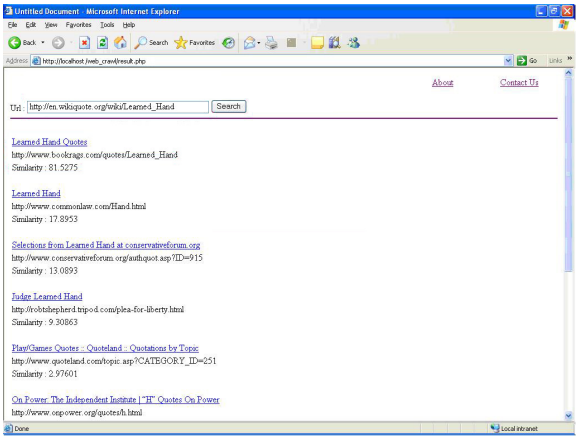


Figure 7. User Search Result Interface

b. Pages for the Administrator:
On the administrator page (Figure 8), the admin can set the number of links in the search, the depth of crawling, what percentage of content in the English language for the web to be processed, the minimum similarity that is displayed, whether the position of words included in the calculations, and limits the location of a words included in the initial word in a paragraph. After doing all the settings, the administrator can enable the crawler and the similarity calculation process. Crawling and also the calculation results of similarity can be seen in Figure 9 and Figure 10.

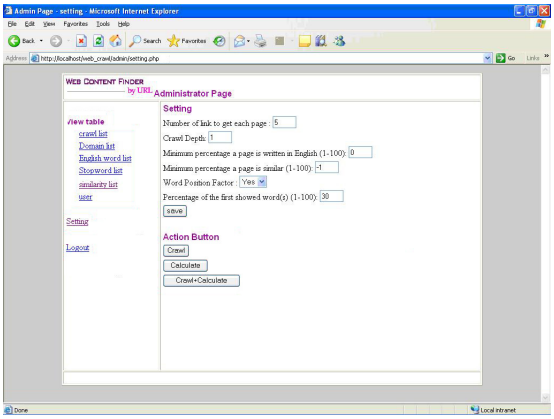


Figure 8. Administrator Setting and Processing Interface

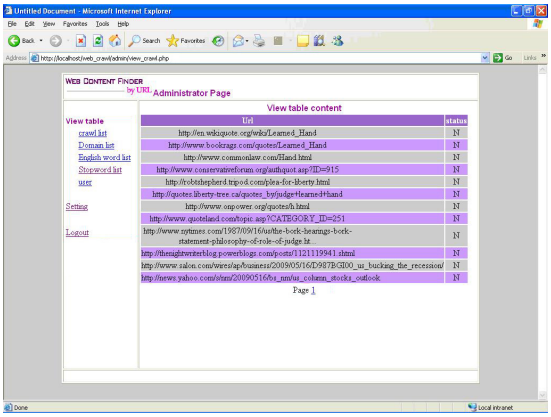


Figure 9. Administrator Crawling Result Interface

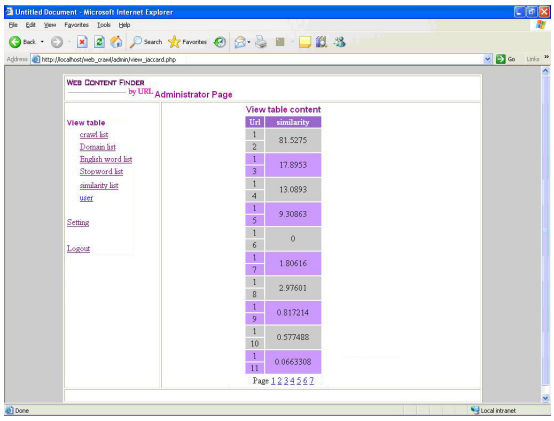


Figure 10. Similarity between Documents Result Interface

5. TESTING

There are two kinds of experiments are performed, namely:

1. Testing of similarity ranking results from the application, which compared with the manually ranked by three sources and site Copyscape (http://www.copyscape.com) that offer similar services. List web pages that are tested and test results can be seen in Table 1 and Table 2.

Table 1. List of Web Pages for The Testing

Id	URL
1	http://www.jhedge.com/story/fiction/bridge.htm
2	http://www.geocities.com/Athens/Acropolis/9343/bridge.htm
3	http://www.bebo.com/Chapters.jsp?ChapterId=3695291437&MemberId=3695253583
4	http://www.strangeroad.com/Stories/Stories100.php
5	http://forum.gamenetworks.com/viewtopic.php?f=228&t=443&st=0&sk=t&sd=a&start=495
6	http://chuntian11.blogspot.com/2007_11_01_archive.html
7	http://chuntian11.blogspot.com/2007/11/bridge.html
8	http://forum.gamenetworks.com/viewtopic.php?f=228&t=443&st=0&sk=t&sd=a&start=490
9	http://mattkline.wordpress.com/

Table 2. Similarity Rangking Comparation

Id URL	Similarity Ranking				
	A	B	C	D	E
1*	-	-	-	-	-
2	1	1	1	1	1
3	2	3	2	2	2
4	3	4	3	4	3
5	4	6	7	6	7
6	5	5	5	7	5
7	6	2	4	3	4
8	7	7	6	5	6
9	8	8	8	8	8

Description:

*: URL id of the web that became the reference

A: The order of ranking the results of www.copyscape.com

B: The order of ranking results from Web Content Finder Application

C: The order of ranking respondent 1

D: The order of ranking respondent 2

E: The order of ranking respondent 3

From the test results can be seen that the ranking of the applications are not much different from the ranking produced by Copyscape site. But the result is quite different when compared to the manual ranking compiled by the respondents.

2. Testing the processing time of all calculations against the number of URLs that are processed. The test result can be seen in Figure 11.

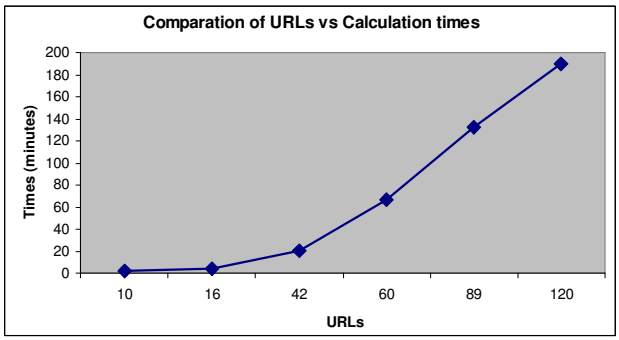


Figure 11. The Comparation of URLs and Processing times

From the test results of the processing time can be concluded that the more URLs that are processed then the longer the process.

6. CONCLUSION

From the comparison of the results of the ranking of URLs can be seen that the system can show good results. Because the results are not much different from the results of the ranking on a professional site that offers similar services. From the calculation speed of the process can be concluded that the application is ready to be implemented.

7. ACKNOWLEDGMENTS

Our thanks to Dr. Rolly Intan, who has helped us to get a better understanding of the research that he has done and also provide any important advice for the research we've done.

8. REFERENCES

[1] Berry, M.W. and Browne, M. 2006. Lecture Notes In Data Mining. World Scientific Publishing Co.

[2] Budhi, G.S., Intan, R., Rostianingsih, S. and Riantarno, S.R. 2007. Indonesian Automated Text Summarization. Proceeding of International Conference on Soft Computing, Intelligent System and Information Technology, Denpasar, Bali.

[3] Chakrabarti, S. 2003. Mining the Web: Discovering Knowledge from Hypertext Data. Morgan Kaufmann Publishers.

[4] Intan, R. and Defeng, A. 2006. HARD: Subject-based Search Engine Menggunakan TF-IDF dan Jaccard's Coefficient. Jurnal Teknik Industri Vol. 8, No. 1.

[5] Porter, M.F. 1980. An algorithm for suffix stripping. Program, 14(3) pp 130-137.

[6] Sjobergh, J. and Araki, K. 2006. Extraction based summarization using a shortest path algorithm. 12th Annual Language Processing Conference NLP2006. Yokohama. Japan.

[7] Turney, P.D. 2000. Learning algorithms for keyphrase extraction. Information Retrieval, 2 (4), 303-336. (NRC #44105).

Developing An Educational Game for 10th Grade Physics Students

Silvia Rostianingsih
Petra Christian University
Siwalakerto 121-131
Surabaya
+62-31-2983455
silvia@petra.ac.id

Gregorius Satia Budhi
Petra Christian University
Siwalakerto 121-131
Surabaya
+62-31-2983455
greg@petra.ac.id

Kestian Olimpik
Petra Christian University
Siwalakerto 121-131
Surabaya

ABSTRACT
An educational game could be used to attract high school students of learning their subjects. This research is focusing on developing a game for 10th grade physics students, especially on Physics Rectilinear and Circular Motion, Newton's Laws of Motion, Geometric Optics, Electrical Dynamics, and Electromagnetic Waves. This game has a storyline on physics lessons. The game is developed using the programming language ActionScript 2.0 and Flash as a local shared object storage media. The end result of this game is a game-based education, which is equipped with the learning board, saving, and loading features. The software is also tested on several different types of computers. The comparison showed that this software can be run well with any standard computer specification.

Keywords
Education Game, Physics, ActionScript

1. INTRODUCTION
The gaming industry is developing rapidly with a variety of genres. The development of the game is influenced by the increasing technological advances resulted in more and more sophisticated games. Based on the results of questionnaires carried out to high school students, it turns out that most of the respondents have difficulties in learning physics because the lack of learning tools. Previous researches have already developed similar games for examples Crayon Physics, World of Goo, and Mifas Adventure; therefore this research combines the favorable games among them.

Features	This game	Crayon Physics	World of Goo	Mifas Adventure
Save game and load game	Yes	Yes	Yes	No
Game explanatory	Clear	Not clear	Not clear	Clear
Degree of difficulty	Average	High	High	Easy
Interface design	Average	Good	Good	Not good

Previous works have already showed the benefit of learning with ICT especially computer games [1] [2].

2. COMPUTER GAME
Computer game is a game played through the aid of computer tools [3]. Computer game is different from other types of games because there is no physical movement or direct interaction with the object except through the medium of the computer. Software must be able to capture the rapid reaction from the interaction between players. The complexity of the game depends on the ability to represent rules and game environment created in the program.

3. PHYSICS THEORY
Some of the theories used in this game are described below.

3.1 Motion
Motion is a change in the position of an object against its surroundings [4] as bellow:

1. Transfer

The position of an object is determined by a line called the axis of coordinates x, where the line is used as a reference point. The value of the displacement of an object depends on the direction of motion. The object displacement is positive if the motion direction moves to the right object. The object displacement is negative if the motion direction moves to the to the right.

2. Velocity

Velocity is a vector which is determined by the magnitude and direction. The velocity of an object is using the following equation:

$$v = s / t \quad (1)$$

Where: v = velocity (m / s); s = displacement (m); t = time (sec)

3. Acceleration

Acceleration is a vector which is determined by the value addition. It is also determined by the direction of displacement. The magnitude of the acceleration equation is using the following equation:

$$a = v / t \quad (2)$$

Where: a = acceleration (m/s2); v = final velocity (m / s) t = time for the velocity change (sec)

3.2 Newton's Law of Motion
Every object persists in its states **of rest or uniform** motion in a straight line unless it is compelled to change that state by forces impressed on it [4]. This can be formulated:

$$F = m * a \quad (3)$$

Where: a = acceleration (m/s2); F = net force (Newton); m = mass (Kg)

3.3 Geometric Optics

Image formation can occur due to a reflection at the mirror which can be divided into flat mirrors, concave mirrors, and convex mirrors [5].

Flat mirror has a reflectivity of law which the angle of incidence equal to the reflection angle. Therefore the two angles are of the same height. The assumption is h is the height of the object and the high point of the t is t = ½ h. Flat mirror also has properties of virtual image and the object distance is equal to the distance between the image and the mirror

Concave mirror is also called convergent mirror. Concave mirror is divided into 4 rooms in the shadows and object placement, where R is the radius and f is the focus, where f obtained from the half of radius. Forming mirror images can be obtained from the special light in the mirror. The first special beam axis is parallel to the primary and will be reflected through the focus. The second one is towards the focus and reflected parallel to the major axis to find intersection. The third one is a special beam of light towards the radius.

Convex mirror is also called diverging mirror. Just like a concave mirror, convex mirror is also divided into 4 rooms. The intersection of the light creates a shadow, so the virtual image will be scaled up. The distance from the mirror to the object called s, while the shadow to the mirror is called s'. The connection between s, s', and f is 1 / f is 1 / s + 1 / s'.

3.4 Electric Dynamic

Electricity is a flow of charge through a point per second, which can be determined [5]:

$$P = V * I \quad (4)$$

Where: P = power (watt); V = voltage (volt); I = the current (amp)

3.5 Electromagnetic Waves

According to James Clark Maxwell, electromagnetic waves occur due to the changes in the magnetic field may cause an electric field. The amount of electromagnetic wave is formulated as follows [5]:

$$c = f * \lambda \quad (5)$$

Where: c = velocity of propagation ($3x10^8$ m/sec); f = frequency (Hz); λ = wavelength (m)

4. GAME DESIGN

The title of this game is "School Day: Bring Back the Genius", which is an education game with a small town setting. The following sections will explain the game's design.

4.1 Environment

The game is in the form of a world map. There are pictures of a home, park, school, museum, garden and the menu menu. The games can choose any of the building to start playing. The menu button consists of choices to Resume, Save, Back to Menu, or Exit. Home is the place where the game begins and the main character gets a mission to accomplish. In this story, home is where the main character lives with his mother. School is the place where the main character can learn, talk to the teacher and her friends (Steven and Cindy), and also can read theoretical physics when talking with the teacher. A garden is the place managed by Toni (the gardener). The main character and his friends can get extra lessons from their teachers in the form of a mini game. Nevertheless this garden also contains missions to be completed. Museum is the place where the main character can see histories of the physicist which is guarded by a security guard. The main character and his friends can get extra lessons from their teachers in the form of a mini game. The flowchart of the game can be seen in Figure 1.

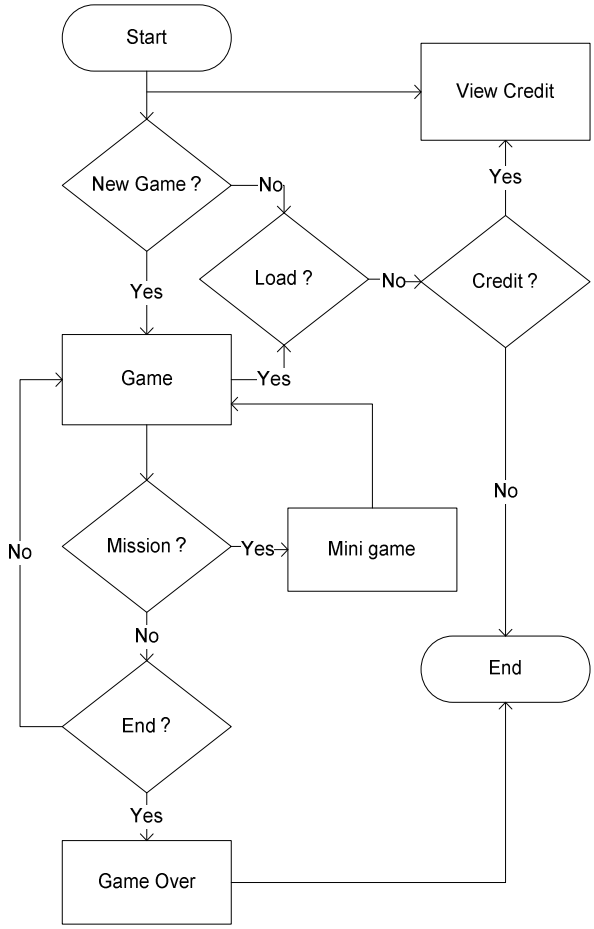


Figure 1. Game Flowchart

4.2 Characters

The characters of the game consist of:

1. Main character (Doan)

Doan is the main character in this game. He is a genius but he wastes his time with playing during his junior high school years. Therefore people around him assume that Doan is not responsible. Doan wants to prove that he is not like what other people think. Therefore Doan wants to change his behavior when he enters senior high school.

2. Rival character (Steven)

Steven is one of the main character's classmates, who has a passion and desire to learn. Since junior high school, he is always top ranking student in school. Therefore he does not like making friends with his classmates who are considered lazy (one of them is Doan).

3. Friend (Cindy)

Cindy is one of the main character's classmates; Cindy is a person who likes to help her friends and has a high enthusiasm for learning. Cindy is not smart as Steven, but she is liked by her classmates.

4. The teacher

The teacher is a teacher who teaches Physics at the main character's class and loves all the students.

5. The gardener (Toni)

Toni is a gardener who is 30 years old. Toni has a desire to develop his garden so that much greater.

6. Security (Erick)

Erick is a guard at the museum who is 22 years old who likes to hang out and learn about history.

7. Mom

The parent of the main character who always gives advice to her son, and hopes his son could change his lazy.

4.3 Stories

"School Day: Bring Back the genius" is an educational game that tells about a boy named Doan who lives in a small town. This story begins on the first day of school when Doan is too late to get up to go to school. When he arrives in school, the teacher gives an introduction and explanation of the motion lesson and illustrates with mini game format (Figure 2). The next day the teacher chooses the president of the class and provides an instruction from the principal to all students to conduct community service in the dirty school environment. Figure 3 represents a game played by Doan when he assigns at the warehouse and surprises by a mouse. Doan and friends are trying to catch the mouse. This game is teaching about the physics lesson on motion. After cleaning the warehouse pack, the teacher explains the Newton's Law of Motion lesson. The teacher advises the students to do extra studies with choices go to the garden (Figure 4) or go to the museums (Figure 5) to replace the time spent for cleaning the school.

The next day, the teacher gives an explanation of geometric optics lesson. Doan must be working on an assignment, which represent by Figure 6.

The teacher will announce the result in the next day. If Doan get more than 80 points then he will be the student who has the highest points. If the score under 80 points, the teacher will advise Doan to study harder. That day, the teacher gives a lesson about electric dynamic. After school, Doan's friends invite him to study together in the park and he must do the electric power mini game.

The next day, the teacher teaches electromagnetic waves lesson by asking questions in class (Figure 7).

The next day is the announcement day where the game is ended. There are 3 kinds of ending point: more than 80 (good), more than 60 and less than 80 (average), and less than 60 (poor).

5. GAME IMPLEMENTATION

5.1 Motion Mini Game 1

The player must answer the speed of the bus (motion) by observing the animation of a walk around the city bus with a speed and distance. If the player answers correctly then he will get 20 points; whereas if the user answers to a range of -2 and +2 correct answer he will get 10 points, otherwise the player get no point. The mini game in Figure 2 uses equation 1. With time is 135 s and distance 8000 m, than v is 59.259 (round to 59).

5.2 Motion Mini Game 2

The player is given an animation of a rat that ran with the speed and erratic. The player should be able to catch the mice with estimating the speed of the mice (motion). If the user is able to guess correctly then he will get 20 points, if the user guessed the wrong answer before time runs out he will get 5 points, otherwise he will get no point. The mini game in Figure 3 uses equation 1, where v = 8 m/s and t = 22, so it results s = 176 m.

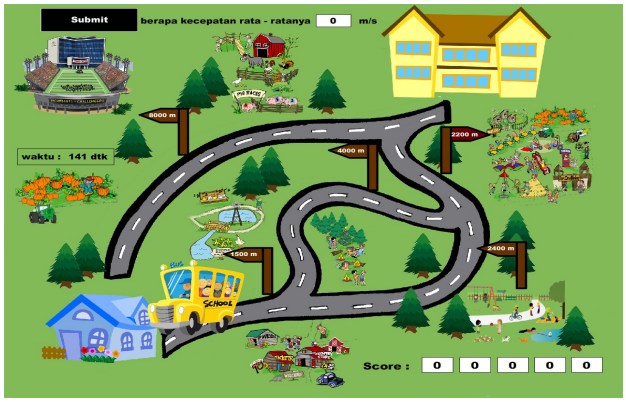


Figure 2. Motion mini game 1

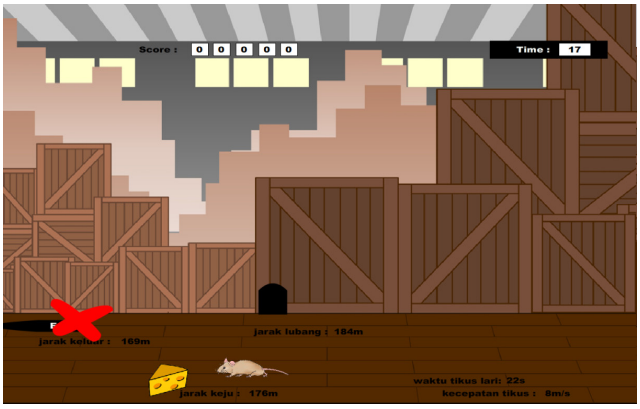


Figure 3. Motion mini game 2

5.3 Newton's Law of Motion Mini Game 1

The main character is pushing an object with a certain weight and fields that the user must press a key / mouse to determine the thrust force (Newton's law of motion). If the user managed to push

the object with a range less than 1 m and -1 m more than the destination then he will get 20 points. If the range is less than 2 m and more than -2 m then get a value of 15, less than 3 m and more than -3 m get 10 points, less than 4 m and more then -4 m get 5 points, else gets no point. The mini game can be seen in Figure 4.

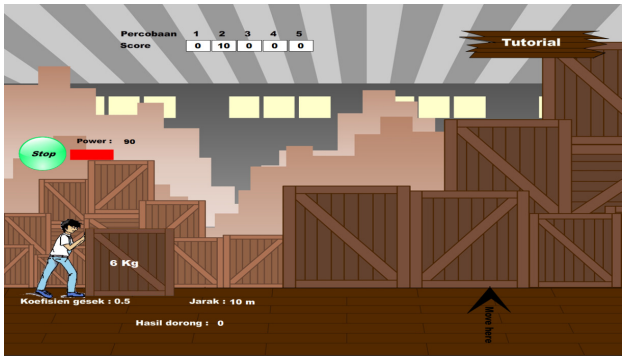


Figure 4. Newton's law of motion mini game 1

Equations uses for this mini game are w (weight) = m (massa) * g (gravity), Fg (friction) = μ (coefficient of friction) * N, s = v0 * t + (1/2) * a * t^2, W = F * s, F = F – Fg, and F = M * a. The step will be w = m * g (6 * 9.8) = 58.8, Fg = μ * N (0.5 * 58.8) = 29.4. The next step is find a with F - Fg = m * a (90 - 29.4 = 6 * a), a is resulted 10.1. s = v0 * t + (1/2) * a * t^2, for v0 = 0 and t = 1 then s = (1/2) * a * t^2, s is resulted 5.05 (round to 5 m).

5.4 Newton's Law of Motion Mini Game 2

The player must be able to estimate the braking acting on the plane with examine the certain mass, speed, and distance of the plane (Newton's laws of motion). If the user managed to land on the runway then the player will get 20 points. If the player landed within range of 10 m, the player will get 16 points, if the range within 20 m the player will get 10 points, and otherwise the player gets no point. The mini game can be seen in Figure 5.

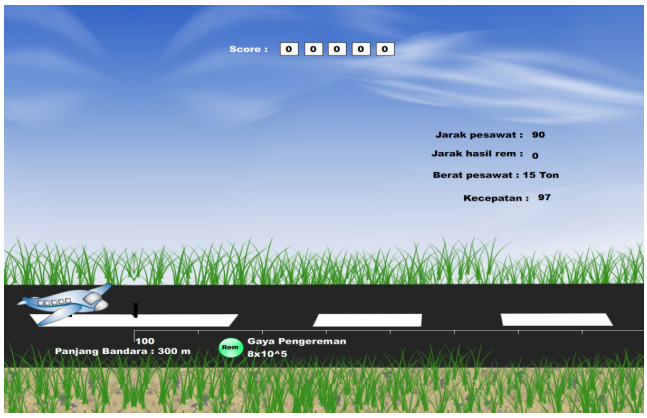


Figure 5. Newton's law of motion mini game 2

This mini game also uses equation vt2 = v02 - 2as (vt = 0) in order to find s. For m = 15 ton, F= 8 * 105, v0= 97, then a = 800000 / 15000 (F = m * a) = 53.33. S = v02 / 2a (s = 972 / 53.33 * 2) = 88,214 (round to 90 m). Total distance will be 90 (initial distance) + 90 (braking distance) = 180.

5.5 Geometric Optics Mini Game

Given an animation of an object with a certain height and distance with a mirror in the middle, there is an object that also has some distance to the mirror and the user must calculate the distance of the shadow of the second object when viewed from a height on the first object (geometric optics). This game has 5 questions. If the player answered each question correctly then the player will get 20 points, if the range of 3 to -3 difference with the correct answer then the user will get 10 points, if the user answered outside the range that has been provided then the player will not get no point. The mini game can be seen in Figure 6.

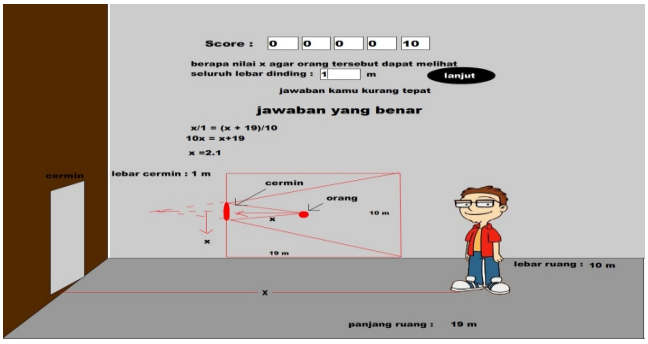


Figure 6. Geometric optics mini game

5.6 Electric Dynamic Mini Game

There is a room which is equipped with electrical appliances. This mini game uses equation 4. The player should switch on/off for each appliance in order to fulfill the potential switch. The maximum score will be 100 points, but when there are appliances which should not be switched, than the score will be reduced. Figure 7 shows the V is 500 volt and the I is 5 amperes, it means the room can only use 2500 watt power.

5.7 Electromagnetic Waves Mini Game

A radio station has a specific wavelength and the player must calculate the right frequency (electromagnetic waves) to receive the signal. This mini game uses equation 5. The player must determine the right frequency before the time runs out five times. Every time the player answered correctly then he will get 20 points, else he will get no points. The mini game can be seen in Figure 8. The mini game shows the wavelength is between 273 m and 373 m while c is $3*10^8$ m/sec. The frequency at 273 m is 1098.901 khz while the frequency at 373 m is 804.289 khz. The right frequency for the radio station is 804.289 - 1098.901 khz.

6. CONCLUSIONS

Application is capable to model physics in the form of game. This game is able to attract high school student to study the whole material by finishing the story. Save and load process in this application uses local shared object.

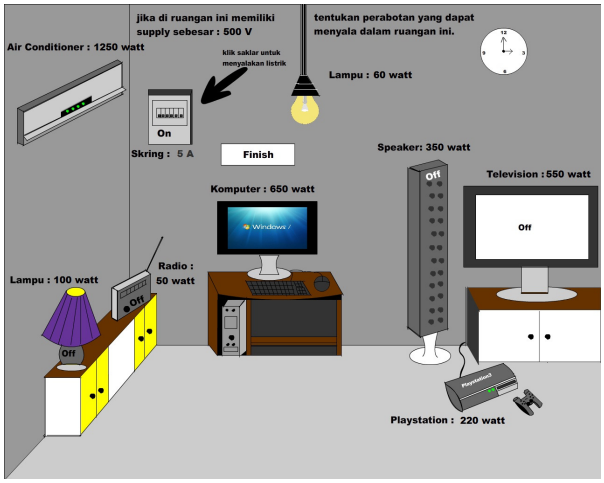


Figure 7. Electric dynamic mini game

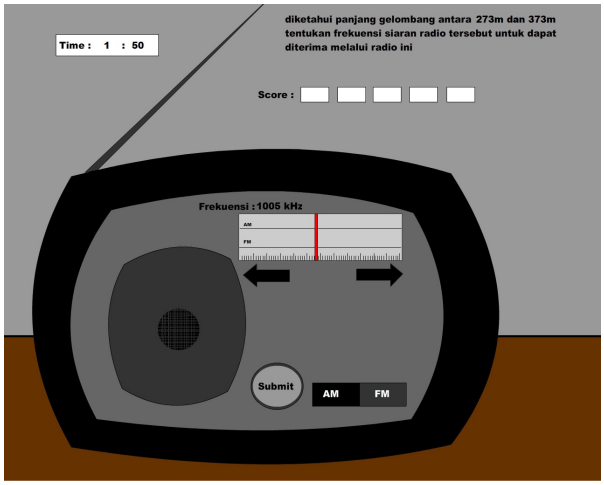


Figure 8. Electromagnetic waves mini game

7. REFERENCES

[1] Tüzün, H., et al. The effects of computer games on primary school students' achievement and motivation in geography learning. Computers & Education (2008). DOI:10.1016/j.compedu.2008.06.008

[2] Papastergiou, M. Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation. Computer 7 Education (2008). DOI: 10.1016/j.compedu.2008.06.004

[3] Prayudi. 2008. Computer game definition. Retrieved February 20, 2009. DOI: http://cgrg.wordpress.com/ 2008/03/12/definisi-computer-game/.

[4] Rosyid, M. F., Resmiyanto, R., Budhi, R. H. S. & Prastya, D. S. B. 2008. Concept Study of Physics 1. Jakarta, PT Tiga Serangkai Pustaka Mandiri.

[5] Tamrin, A. J. 2009. Smart for high school 10, 11, 12 grades in Physics. Surabaya, Gitamedia Press.

Design Enterprise Architecture using E2AF for Retail Company

Lily Puspa Dewi
Informatics Dept. Petra Christian University
Siwalankerto 121-131
Surabaya, Indonesia
62-31-8439040
lily@petra.ac.id

Uce Indahyanti
Graduate Student
+628123155056
u.indahyanti@gmail.com

Yulius Hari
Graduate Student
+886 - 975969548
yulius.hari.s@gmail.com

ABSTRACT
Utilization of information technology (IT) provides many advantages in all aspects of the business, which include providing on accurate, rapid and integrated to support all business functions. Use of IT in retail is needed primarily to support the supply chain network (supply chain network) and improve on customer service. This paper tries to discuss the design of Enterprise Architecture (EA) on the retail company business, and as a reference in the design, we use the Extended Enterprise Architecture Framework (E2AF).

The results are EA2F can give the retail company a big picture about "the organization-wide roadmap" for achieving the vision and mission of the organization, through optimal performance of core business processes within an efficient information technology (IT). The design of EA used to map the entire roadmap E2AF Company including its style, is a strategic step that is necessary to achieve the vision and mission. The design of this EA is still a map of “As Is” condition that needs to be addressed further in order to be perfectly deliverable.

Keywords
Enterprise Architecture, E2AF, retail company, Supply Chain

1. INTRODUCTION
In line with the development of information technology in Indonesia, almost all of our activities can not be separated from the touch of technology, ranging from information service provider to our work facilitator and not least in the retail company. Retail Company has never missed with the use of technology as a information media, technology now has an important role in the retail company, such as an information system that is able to accommodate the supply chain between the supplier or warehouse stores. Similarly, as a media service that is able to provide service and communication between the company and customer.

PT ABC is a one company in the Retail franchising which growing rapidly in Indonesia today. At the end of 2009, it already has 3293 outlets spread across the island of Java and Bali. In the same year, it becomes a public company in Indonesia Stock Exchange.

PT ABC also runs several systems that have been integrated such as a communication media with its customers, as well as with the franchise business owners, as well as stakeholders and suppliers. Besides that, the focus also to the stock availability that is always kept the quantity. PT ABC does not want to disappoint customers because lack of stocks where in addition to the lost sales are also able to create a bad impression or make disloyal customer.

2. UNIQUE CHARACTERISTIC OF RETAIL COMPANY
The uniqueness of the retail company system, among others, is:

1. No manufacturing processes

   Retail Company does not do manufacturing process. The retail company has a high dependence on suppliers, because it is only authorized as a wholesale of existing products and it follows the policy of selling products from the Supplier.

2. Good distribution network

   Retail distribution channels consist of some combination of producers or manufacturers, agents or brokers, wholesalers or distributors, importers, and retailers. Each step along the channel has a specific purpose that is met by one or more member companies. Distribution channels are important because they allow for a continuous flow of product despite the natural peaks and slumps experienced in manufacturing and sales. They also provide efficiency, economies of scale, and cost savings to members of the channel.

3. Price stability

   One interesting point is the uniqueness of every retail business has a uniform price for both at the central and branch. And generally every retail company to compete and compete in the price.

4. Centralized warehouse and supplier

   One of the things that support the retail company has a centralized distribution system supplier which already has its own warehouse to distribute the goods to the branches. Hence, in the order to the supplier can be done in a very large quantity so as to depress prices.

5. Approach to customer behavior

   In retail enterprise is not just a personal approach, but also pay attention to consumer behavior, in addition to seeing sales data which are often purchased, various promotions are also conducted based on his behavior.

The examples are frequently purchased items will be placed on the middle shelf is easily accessible by consumers. Retail business is also competing to hold a promo or event based on a national holiday.

3. BUSINESS PROCESS AND DISTRIBUTION

In order to guarantee the availability of various products to thousands of customers' everyday, as well as creating efficiencies for itself and its suppliers, PT ABC reorganize their supply chain system. It is important to maintain the availability of stock, especially because the product being sold is grocery.

Business processes for supply chain in this retail company uses the Just In Time (JIT) in the Distribution Center (DC) called Cross Dock. The purpose of this business process is to reduce the goods lead time before distribute to the outlets. So, when the supplier sends the goods to DC today, then the next day that the goods have been sent to the outlets. In short, the method of Cross Dock allows the process more transparent in the distribution of the products because there is no degradation product (left) in the warehouse. Basically, the function of DC is to redistribute the product, not to store the product.

The uniqueness of the way – comparing than the supplier sends directly - that the products had been consolidated when delivered to stores. For example, normally, grocery outlet receives 30 different supplier's trucks, because, the supplier may submit to the DC. Furthermore, the goods from various suppliers will be sorted out according to demand outlets. For example, now a truck that came to the outlet just need to bring products that are specifically required by that outlet

The developed supply chain is not only based on the physical movement of products, but also to the flow of information. The success in the retail supply chain is largely determined by the flow of information from the outlet to the supplier, and vice versa, data synchronization with both parties.

For order process flow, Retail Company developed model as Central Orders Pool (COP) in which the ordering process is done automatically and centrally based on the position of stock in outlet and other parameters. To make purchase orders to all suppliers, Retail Company uses an Electronic Data Interchange (EDI) to collect products list that will be ordered from every outlet. The product list will be announced over the web, so supplier can check what products they should send.

Because the purchase order process is centralized, the accuracy stock amount in every outlet is important. To maintain this accuracy, the retail company adjusts with cycle count process that counts the stock every day. Hopefully, data accuracy in distribution center is valid although they handle a thousand products.

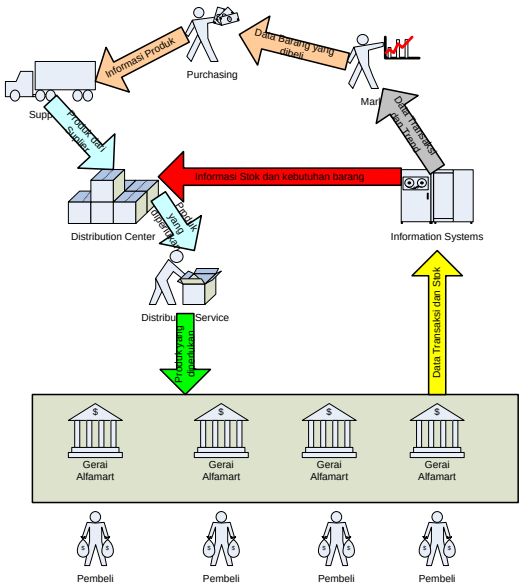


Figure 1. Business Process and Distribution

Every outlet which incorporated with this retail company integrate in to same information system as shown in Figure 1.

The system will automatically take advantage of the concept of Just In Time (JIT) so that the groceries that be required having minimal stock. When the products have reached the minimum stock then the system will automatically order that product to Distribution Center (DC) to be forwarded to the distribution division to deliver the product to the corresponding outlets. Sales transaction data can also be used by marketing division to determine strategy or next purchase order. The purchase order will be handled by purchasing division to be submitted to supplier.

4. EXTENDED ENTERPRISE ARCHITECTURE FRAMEWORK (E2AF)

Institute For Enterprise Architecture Developments (IFEAD) has developed architectural design methods, which prescribe a coherent design and realization of new business and the supporting IT systems. This guarantees the full integration between the business and human perspective of an organization and the technology functionality of supporting IT systems (Institute for Architecture Enterprise Developments)

IFEAD describes architecture as a set of principles, rules, standards and guidelines reflecting the organizational culture and behavior that prescribe architects, program/program managers and developers how to deal with the transformation of both the business and IT systems.

4.1 Definition

IFEAD uses the following definition for enterprise architecture: Enterprise Architecture is about understanding all of the different elements that go to make up the Enterprise and how those elements interrelate. Enterprise Architecture embodies a set of principles, rules, standards and guidelines, expressing and visualizing the vision, culture and behavior of an organization while implementing certain concepts that serves as prescription

for the design and construction of a certain object type. It contains a combination of style, engineering and construction principles, guaranteeing the uniformity and quality of the resulting object.

IFEAD has developed such an architectural approach for the design and realization of both the business & Information areas of an organization as well as for the supporting IT systems. This approach is applicable for different organizations, in different situations and at different contemplation levels.

The architecture style reflects the philosophy and mindset behind the framework and approaches and delivers a certain commonality in execution with respect to organizations unique situation.

4.2 The role of Architecture

In the development of a house, building or any object we can always identify the following main steps (Figure 2):

- A discovery process to identify the needs and requirements in the context of a certain situation;
- A design process which leads to a design of the object in the form of drawings and/or models;
- A transformation process to plan the realization of the object in its environment;
- A construction process that regards the realisation of the actual object based on the design and realization plan.

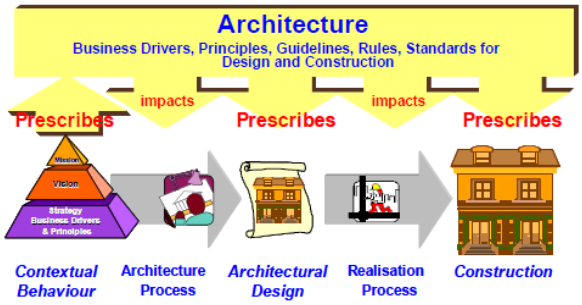


Figure 2. Development Steps

The principles, guidelines and rules identified in the discovery phase are used in both the design, transformation and construction process. As such, the architecture impacts all processes. The architecture constraints the freedom of the designer and constructor of the object and guides them towards a structure that complies with the business vision and concepts of the architecture. The architecture serves as a prescription for the design, transformation and construction of the object. As a result the object will be recognised as being 'designed and constructed under architecture'.

The three main elements in E2AF are:

- Construction
- Function
- Style that show company's culture, value, and rules.

The term enterprise architecture associated with the construction and function, while the style is forgotten or ignored. With three main elements of the above, demonstrate an understanding of enterprise architecture differences among these elements and how these elements interrelate to support the organization.

E2AF's excellence can make sure the enterprise designer take full advantages of the alignment between business and IT by integrating all areas of architecture into a single whole result. Architecture design consists of business, information, information systems, security, infrastructure, and governance aspects. The design will become like a map for the management to organize all relevant matters.

E2AF based on the ideas described in IEEE 1471-2000 about viewpoint and the transformation from concepts into the enterprise architecture. Some companies' stakeholders influence on the goals, objectives and characteristics. In addition, the stakeholder also has a different concern and even a different viewpoint. The model of stakeholder is divided into four types, such as Business, Information, System and Technology Infrastructure as shown in the below.

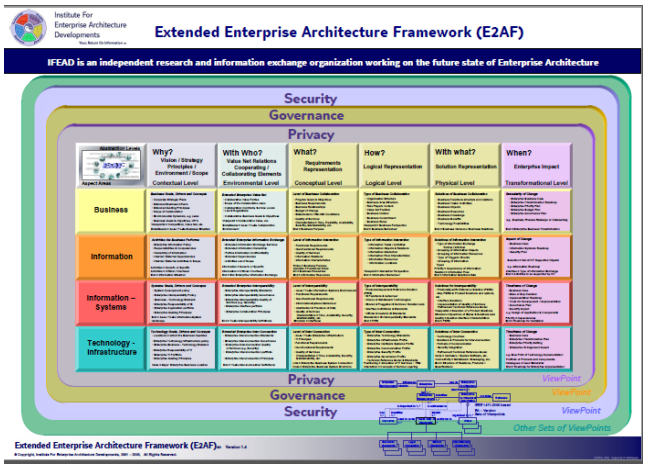


Figure 3. Extended Enterprise Architecture Framework (Institute For Enterprise Architecture Developments)

The topics are divided into:

- Contextual Level (Why) describe the context and scope of the company that is more profound. "Why" describe the motivation of the company, a statement about the vision, mission, scope and business strategy and technology firm.
- Environmental level (With Who) describe the business relationship and how the flow of information. "WHO" describes the relationship with business and technology in the extended (enterprise)
- Conceptual level (What) illustrate the need. "What" describes the purpose and objectivity as well as the needs of all corporate entities involved in all aspects.
- Logical level (How) describes the ideal solutions
- Physical level (With what) describe the products and technical solutions physically. "With what" describe physical solutions in every area, including changes in

business and communications, software and tools, hardware and communication products.

- Transformation level (When) describes the impact of a solution to the company. When describing the transformation roadmap, on the whole area.

5. ENTERPRISE ARCHITECTURE DESIGN

Enterprise Architect (EA) provides a map of the entire organization (the organization wide roadmap) to achieving the vision and mission of the organization, through optimal performance of core business processes within an efficient information technology (IT). Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

EA is also integrated in the Information Architecture (AI) which is a comprehensive model of the data, business processes and IT assets within the company. AI presents a long-term view of the various processes, systems and technologies based on a consistent and coherent design.

In designing the EA really need attention to the details of any existing business processes within the company, in order to achieve harmony between the structures of the company's IT with business goals.

Company's business processes have been discussed above, and, we summarized in the following few points:

- All outlets integrated into an information system which capable of displaying transaction data and the condition of stocks of goods.
- Supply chain business processes using the Just In Time (JIT) in the distribution center or Distribution Center (DC) called Cross Dock.
- The Cross Dock does not make a degradation of products (left) in the warehouse, because the function of DC is to redistribute the product, instead of storing the product
- The system will automatically do a purchase order to the Distribution Center (DC), if one product has reached the minimum stock
- Ordering to all suppliers uses an Electronic Data Interchange (EDI), a supplier can receive it via Web. Some supplier has been integrated with their ERP system.
- Retail company implement cycle count process (stock checking using daily sampling). That way, the accuracy of the data in the DC claimed almost always 100%, while managing 10.000 of products
- DC will distribute certain goods to outlets accordance with the order.
- Sales transaction data will be used by marketing parties to determine strategy and forecasting purchase order, so the stock amount in DC is always fulfilled.

6. E2AF MAPPING

The design that has "the organization-wide roadmap" we lay out below. The E2AF is used to map the retail company's enterprise architecture.

As shown in Figure 4, the business roadmap was mapped into six levels.

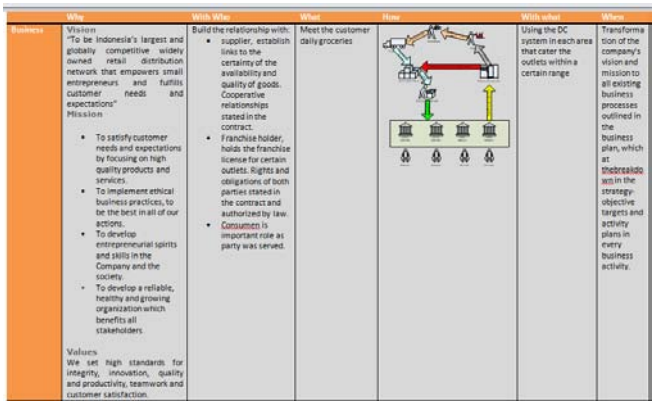


Figure 4. Business' Organization Roadmap

In Figure 5, the Information of Retail Company was mapped. This will be given a clear view of information flow.

Figure 5. Information's Organization Roadmap

In Figure 6 shown retail company information system in every level.

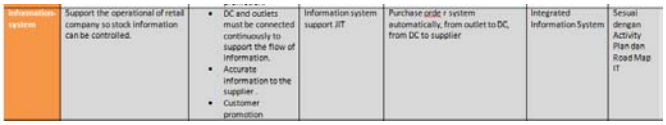


Figure 6. Information System's Organization Roadmap

In Figure 7 shown technology infrastructure support the company.

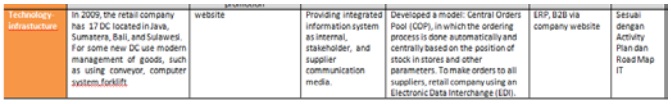


Figure 7. Technology Infrastructure's Organization Roadmap

7. CONCLUSION

- The design of EA set "the organization-wide roadmap" for achieving the vision and mission of the organization, through optimal performance of core business processes within an efficient information technology (IT)

- The retail company has some unique characteristics, the company focuses its business processes in to the flow of products (goods) from supplier to customer
- The design of EA used to map the entire roadmap E2AF Company including its style, is a strategic step that is necessary to achieve the vision and mission.
- The design of this EA is still a map of "As Is" condition that needs to be addressed further in order to be perfectly deliverable.

8. REFERENCES

[1] Insitute For Enterprise Architecture Developments, 2006. E2AF A0 New Poster 03-2005 version 1.4. (Dec. 2011). http://www.enterprise-architecture.info/Images/E2AF/E2AF%20A0%20New%20Poster%2003-2005%20version%201.4.pdf

[2] Institute For Enterprise Architecture Developments, 2006. Extended Enterprise Architecture Development Essential Guide. (Dec. 2011). http://www.enterprise-architecture.info/Images/E2AF/E2AF%20Essentials%20Guide%20%20Description%20v48.pdf

[3] Javadi, M. M. 2006. Extended Enterprise Architecture Framework E2AF (Nov. 2011). http://www.cioindex.com/CIO_Whitepapers/View_CIO_Whitepaper/ArticleId/49525/Extended-Enterprise-Architecture-Framework-E2AF.aspx

[4] Minoli, D. 2008. Enterprise Architecture A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology. Auerbach Publications.

Application of Multi Criteria Decision Making for an Online Awardees Short Listing System

Leo Willyanto Santoso
Informatics Department
Petra Christian University
Jl. Siwalankerto 121-131
Surabaya, INDONESIA
leow@petra.ac.id

Lukas F. Kaiwai
Informatics Department
Petra Christian University
Jl. Siwalankerto 121-131
Surabaya, INDONESIA

Alexander Setiawan
Informatics Department
Petra Christian University
Jl. Siwalankerto 121-131
Surabaya, INDONESIA
alexander@petra.ac.id

ABSTRACT
The selecting candidate is a usual process in any organization, for example personnel recruitment, supplier selection, and awardees short listing system. The selection process in the area of awarding scholarships often uses a manual process which examines the candidate's qualifications according to that specified by the scholarship scheme. The main objective of this research is to reduce the inefficiencies within the process and improving the quality of the final decision by minimizing the degree of personal preference. In this research, Analytical Hierarchy Process (AHP) has been implemented in awardees short listing system at Petra Christian University. Based on these needs, required a system that can provide considerations determine the decision of the scholarship recipients. Such systems provide objective results based on the terms and conditions have been determined by comparing the value of each scholarship recipients. So the result is in accordance with input provided by potential recipients. Based on the questionnaire result, this application meets with end user requirements, 70% of the respondents said this application improve the efficiency of decision making process in terms of multi criteria decision making ability.

Keywords
Analytical Hierarchy Process, Scholarship, Decision Making.

1. INTRODUCTION
Petra Christian University Surabaya is one of the oldest private universities in Surabaya. By continually improving the quality of education, Petra Christian University continues to improve services and facilities that support academic atmosphere, such as by providing scholarships for students. A scholarship is an award of financial aid for a student to further their education. Scholarships are awarded on various criteria, which usually reflect the values and purposes of the donor or founder of the award. The types of scholarships are: merit-based scholarship, need-based scholarship, and student specific scholarship.

Scholarships are offered not only for undergraduate students, but for prospective students as well. In the awardees short listing system, Petra Christian University requires an application that can assist in determining priorities which candidate will receive a scholarship, so that decisions can be made efficiently. It requires a method that can assist in determining priorities so as to assist in decision making.

One method of decision making that can be used in the selection process of scholarship recipients is the method of Analytical Hierarchy Process (AHP). AHP is a systematic method and the appropriate selection using the method of decision-making based on a systematic analysis of the data.

The purpose of this research is to develop an awardees short listing application to support decision making in the selection of candidates for scholarship recipients using AHP in accordance with criteria set by Petra Christian University Surabaya.

The remaining part of this paper is organized as follows. Section 2 presents an overview of current proposal for dealing with analytical hierarchy process. Section 3 presents the methodology of this research and the approach that we have delineated to solve the proposed problems. Section 4 discusses the performance of proposed methods. Finally, section 5 concludes the paper.

2. ANALYTICAL HIERARCHY PROCESS
AHP is one of new approach to solve Multi-Criteria Decision Making (MCDM) problem which reduces complicated criteria to a series of pair wise comparisons and synthesizes the results [1, 2, 3]. One of the real world problems which could be encountered by MCDM is the recruitment and selection processes in human resources department. AHP have been implemented extensively in the recruitment and selection of human resources. Kaka [4], Katsumura [5], Melon [6] and Dolan [7] implemented AHP in the area of pricing system selection, cancer screening option, educational project evaluation, and patient preferences respectively.

AHP can be combined with other technique, like fuzzy system. This combination could solve problems in supplier selection [8], maintenance decision [9], organizational performance [10], human resource evaluation [11] and terminal selection [12].

AHP was chosen for this study based on the following reasons:
- The ability of AHP to incorporate tangible and intangible factors in a systematic way.
- It can solve constructed problems in a variety of decision making situation, ranging from the simple personal decisions to the complex capital intensive decision.
- The problem is broken down in a logical fashion from the large elements to smaller elements.
- It works by examining judgments made by decision makers and measure the consistently of those judgments.

AHP define the frameworks with a hierarchy of objective. It helps the user to achieve the goal from the attributes of problems, by

decompose into a hierarchy. A hierarchy is a three-like structure that represents a complex problem on a number of levels [2]. The hierarchy develop consists of several levels. The top level represents the goal or the objective in terms of a problem statement. At the next level, the major criteria are defined is broad terms. Each criterion may be broken down to smaller level or individual parameter depending on the how many details are called in the model. The bottom level of the hierarchy contains the actual condition of the alternatives which under laying the problem.

Once the hierarchy has been structured, the next step is to establish the priority each factor on a given level of the hierarchy. The decision-maker makes judgments about the relative importance of the element with respect to elements at the higher level in the hierarchy using pairwise comparison. In the AHP pairwise comparison a nine-point scale. (1-9) ratio be used to quantify the decision make's strength of feeling between any two alternatives with respect to given criterion.

In the pairwise comparison a matrix is the preferred form. In general, if 'n' (elements) are being compared for given set of matrix, a total of n(n-1)/2 judgments are necessary to fill in the matrix. Saaty (1995) describes pairwise comparison on matrix as the element that appears in the left hand column is always compared with the element appearing in the top row, and the value is given to the element in the column as it compared with the element in the row. The next process is synthesis.

Synthesis is a process to complete the relative weights of the elements with regard to an element on the next level. This process must be performed for all matrix developed in the pairwise comparison, by using normalized eigenvector associated with the longest eigenvalue of their comparison matrix.

One of the advantages of the AHP is that it provides consistency checking of judgments call consistency Index (CI). Consistency is the intensities of relations among ideas or objects based on a particular criterion justify each other in same logical way. To add perspective, the consistency index can be verified in terms of inconsistency ratio (IR). Experience suggest that the CR or IR value should be between 0 and 0.10 or within ten percents of what would be the outcome from random judgments is acceptable [2].

3. METHODOLOGY

In this part, research methodology will be described. The methodology involves using selected tools and techniques. There are Microsoft Project 2007 to manage the project, Power Designer 6.0 as a CASE tool, PHP as a programming language and MySQL as a database management system. All these tools are very important to ensure the project is carried out smoothly. Figure 1 presents research methodology.

The framework is developed through System Development Life Cycle (SDLC) methodology. The first step is investigation and planning phase. Literature review and knowledge acquisition are done. In this research, literature review is focus on two main topics, "selection criteria" and "AHP". By conducting questionnaires survey and semi-structured interviews, expert knowledge captured. Figure 2 presents criteria for awardees short listing.

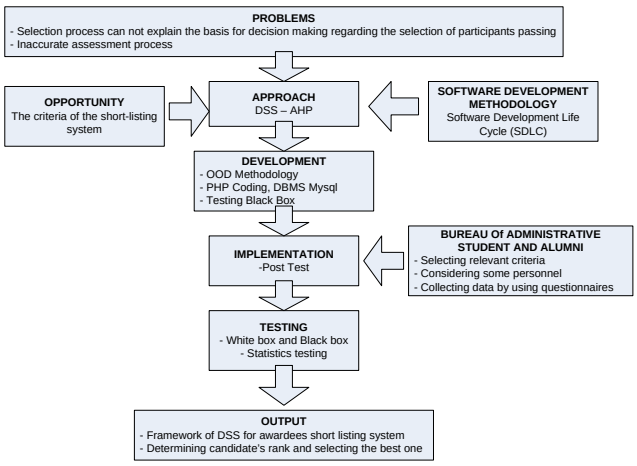
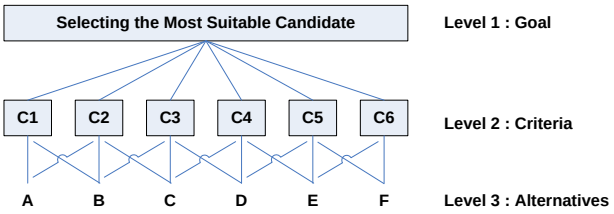


Figure 1. Research Methodology



A, B, C, D, E and F are the applicants being pre-qualified

Criteria	Description
C1	Academic Achievement
C2	Socioeconomic Status
C3	Personality
C4	Leadership Qualities
C5	Cooperative Attitude

Figure 2. Awardees Short Listing System and Criteria

In this research, context diagram and data flow diagram (DFD) are used for modeling processes, while Entity Relationship Diagram (ERD) is used to define the data storage for the system. Figure 3 shows system context and data flow diagram of the system. Figure 4 shows entity relationship diagram of the system.

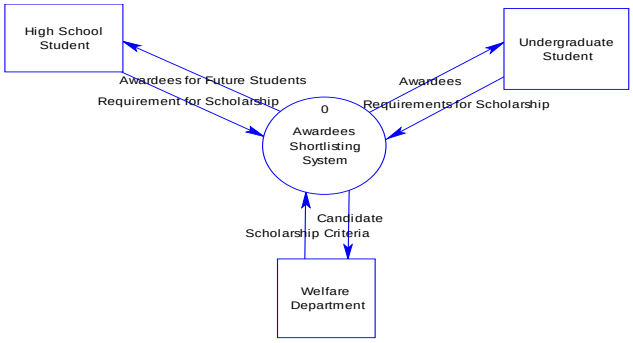


Figure 3. Context Diagram

Application software is developed through process analysis, design, coding, whereby the software will be able to support the implementation of business processes of an organization, such as management of applicants, testing, and supporting decision for awardees short listing system.

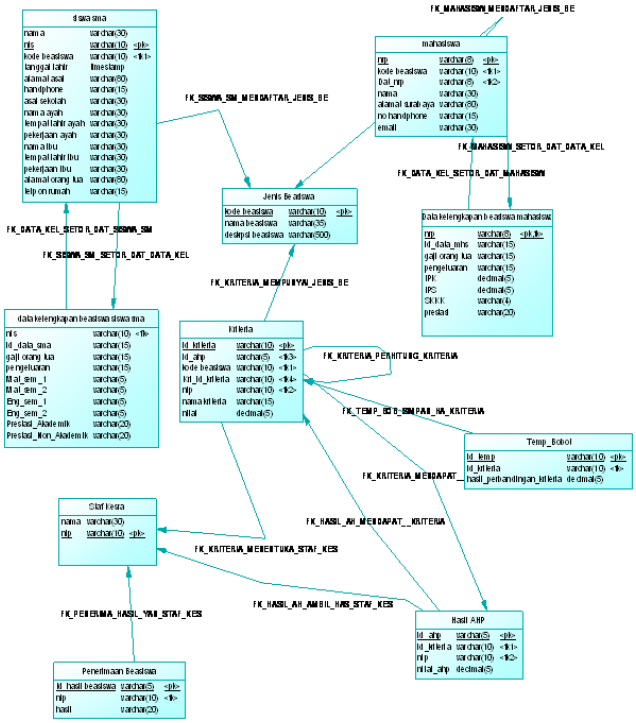


Figure 4. Entity Relationship Diagram

The main sub-system of the proposed application is developing AHP. The design of the AHP must satisfy the goal of developing model that allowed the management to decide which candidate to be selected from the list of qualified candidate, the problems in the bureau of administrative student and alumni (BAKA/Biro Administrasi Kemahasiswaan dan Alumni) are numerous and the interrelationships among the criteria are extremely complicated. In any model development of AHP, the users needs to identify the goals of the research and the problem faced in order to achieve the goals [2].

The procedure for using the AHP can be summarized as [3]

1. Model the problem as a hierarchy containing the decision goal, the alternatives for reaching it, and the criteria for evaluating the alternatives. Fig. 2 shows the criteria which used by the system.
2. Establish priorities among the elements of the hierarchy by making a series of judgments based on pairwise comparisons of the elements. Table 1 shows scale of pairwise comparison. The table is used to compare each criterion with each other criterion, one-by-one.
3. Synthesize these judgments to yield a set of overall priorities for the hierarchy.
4. Check the consistency of the judgments. Experience suggest that the consistency ratio (CR) or inconsistency ratio (IR) value should be between 0 and 0.10 or within ten percents of what would be the outcome from random judgments is acceptable [2].
5. Get a final decision based on the results of this process.

4. DISCUSSION

This part described in detail the testing process for the application that has been implemented. Tests performed on each menu as a whole, with the aim to be seen whether the application runs well. The application was built based on the theoretical model, which has demonstrated how the user systematically can adjust the criteria to improved decision making process. The application has all the capability to evaluate the criteria and stress the intuitive judgment of the decision maker and justified the decision using graphical reports and sensitivity analysis. Figure 5 shows the main page of the application.

Figure 5. Screen Capture of Main Page

Firstly, data of each scholarship applicant entered into the page of applicants. Inputted data is personal data, such as name, address, place and date of birth. Moreover, there is also academic data, which includes academic achievement, education background, and English proficiency test (Fig 6). After that, criteria and sub-criteria inputted to the system, which followed by the score for each criteria and sub criteria for each available scholarship. Figure 7 depicts criteria and sub criteria of the system, and Figure 8 displays inputting value for each criteria and sub criteria. In this system, 5 criteria are used (Fig. 2). This program is built in a flexible, where the criteria can be removed, replaced and added. For example, Academic Achievement (C1) has Sub Criteria like GPA, cumulative GPA, and Awards.

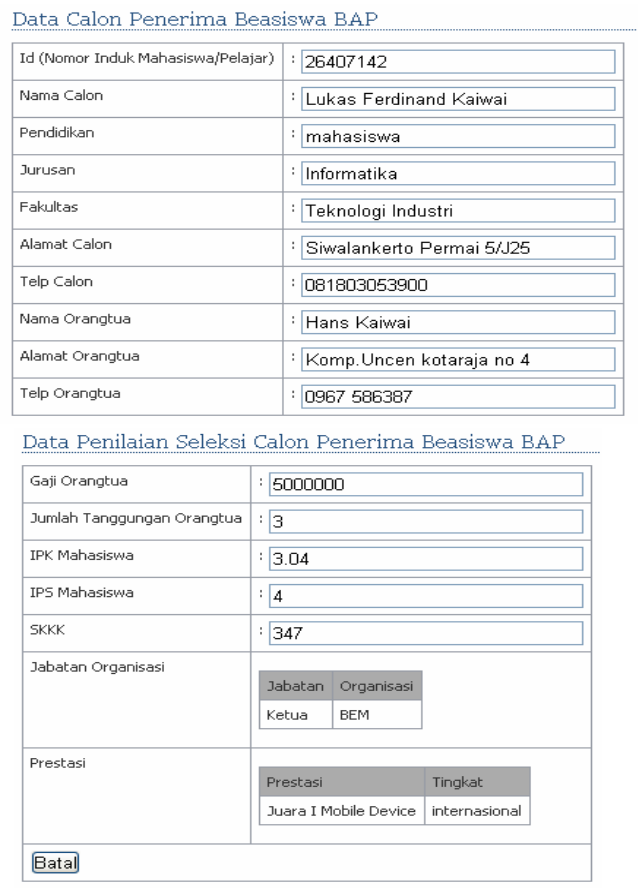


Figure 6. Screen Capture of Entry Page of Applicant

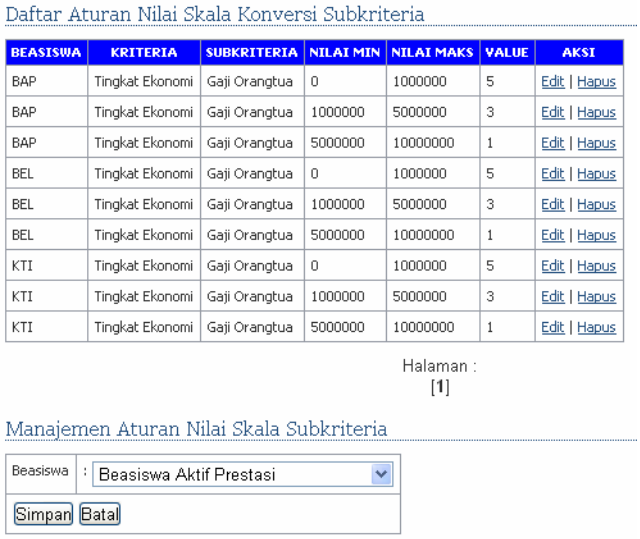


Figure 7. Screen Capture of Criteria for Selection Model

Figure 8. Screen Capture of Comparison Form

This program can be used to conduct an assessment of existing candidate within the organization. The system will process the score of each criteria based on the results of the criteria multiplied by the value of the sub-criteria. The expectation is if there are applicants who are less fit, it is advisable to be transferred to another position. The results of the implementation of the system can be seen in Figure 9 with the results that have been sorted by the highest score of an applicant.

NAMA_CALON	JURUSAN/PRODI	FAKULTAS	TOTAL_NILAI
RW	Ilmu Komunikasi	Ilmu Komunikasi	2.21378
AD	Desain Interior	Seni dan Desain	2.21014
KA	Manajemen Pariwisata	Ekonomi	2.08495
CP	Manajemen Pemasaran	Ekonomi	2.08426
HMT	Manajemen Bisnis	Ekonomi	2.07436
OB	Informatika	Teknologi Industri	2.07384

Figure 9. Screen Capture of Final Candidates

According to this calculation, it can be said that, candidate can be ranked in descending order according to their total score (Fig. 9) as:

- RW, from Faculty of Communication Studies having a score of 2.21378.
- AD, from Faculty of Art and Design having a score of 2.21014.
- KA, from Faculty of Economics having a score of 2.08495.
- CP, from Faculty of Economics having a score of 2.08426.
- HMT, from Faculty of Economics having a score of 2.07436.
- OB, from Faculty of Industrial Technology having a score of 2.07436.

Based on the investigation it was identified that though AHP has an enormous impact in reducing the degree of subjectivity in the decision making process. The Consistency Ratio (CR) was used to inform the decision maker is any personal preference was involved into the recruitment and selection process.

5. CONCLUSIONS

In this paper, AHP approach for awardees short listing system was presented. The method was applied using data from a real case in the educational industries. This application will generate the output score of candidates, so that applicants who have the highest score will have a great opportunity to be awarded a scholarship in the educational industry. Based on the questionnaire, gathered from head of BAKA and Welfare Policy Staff, the results show that:

- 70% of the respondents said that the features of this application are complete, including entering applicant' data until generating the reports.
- 70% of the respondents investigated that this application improves the efficiency of decision making process in terms of multi criteria decision making ability.

For the future research, considering the fuzziness of executives' judgment during the decision-making process is crucial. By using fuzzy theory, it can be implemented to reduce uncertainty condition.

6. REFERENCES

[1] Vaidya, O.S. and Kumar, S. 2006. Analytic hierarchy process: an overview of applications. European Journal of Operation Research, vol. 169, 1-29.

[2] Saaty, T.L. 1980. The Analytic Hierarchy Process, McGraw-Hill International, New York.

[3] Saaty, T.L. 2008. Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World. Pittsburgh, Pennsylvania: RWS Publications.

[4] Kaka, A., Wong, C., Fortune, C., and Langford, D. 2008. Culture change through the use of appropriate pricing systems, Engineering, Construction and Architectural Management, vol. 15(1), 66-77.

[5] Katsumura, Y., Yasunaga, H., Imamura, T., Ohe, K., and Oyama, H. 2008. Relationship between risk information on total colonoscopy and patient preferences for colorectal cancer screening options: Analysis using the Analytic Hierarchy Process, BMC Health Services Research, vol. 8, 106.

[6] Melon, M., Beltran, P., and Cruz , M. 2008. An AHP-based evaluation procedure for Innovative Educational Projects: a face-to-face vs. computer-mediated case study, Omega, vol. 36(5), 754-765(12).

[7] Dolan, J. and Iadarola, S. 2008. Risk communication formats for low probability events: an exploratory study of patient preferences, BMC Medical Informatics and Decision Making, vol. 8(14), pp. 14, April 2008.

[8] Bottani, E. and Rizzi, A. 2008. An adapted multi-criteria approach to suppliers and products selection-An application oriented to lead-time reduction, Int. J. Production Economics, vol. III, 763-781.

[9] Cascales, M. and Lamata, M. 2008. Fuzzy Analytical Hierarchy Process in Maintenance Problem, IEA/AIE, LNAI 5027, 815-824.

[10] Wu, C., Chang, C., and Lin, H. 2008. FAHP Sensitivity Analysis for Measurement Nonprofit Organizational Performance, Quality & Quantity, vol. 42, 283-302, 2008.

[11] Genevois, M. and Albayrak, Y. 2008. A fuzzy multiattribute decision making model to evaluate human resource flexibility problem, Journal of Multiple-Valued Logic and Soft Computing, vol. 14(3) (5), 495-509.

[12] Ma, F., Xu, G., and Yang, F. 2011. Capability adaptation algorithm based on joint network and terminal selection in heterogeneous networks, Journal of China Universities of Posts and Telecommunications, vol. 18 (1), 76-82, September 2011.

Adaptive Information System Life Cycle: Case Study on Petra Christian University Library

Adi Wibowo
Informatics Department
Petra Christian University
Siwalankerto 121-131 Surabaya
+62312983455

adiw@peter.petra.ac.id

ABSTRACT

Petra Christian University Library faces several problems in developing and maintaining its IS portfolio. These problems arised from IS strategy miscalculation and then caused business strategy miscalculations too. To overcome these problems an IS life cycle that fit with PCU library conditions needs to be identified. This research proposes IS life cycle that uses principles such as incremental finding and execution, time boxed development, and knowledge management that also concerned with formation of business and IS strategy.

Keywords

Information system life cycle, business process management.

1. INTRODUCTION

Petra Christian University Library has a software development division. This division analyses, designs, and develops PCU library information system according to the library's needs and strategic plan.

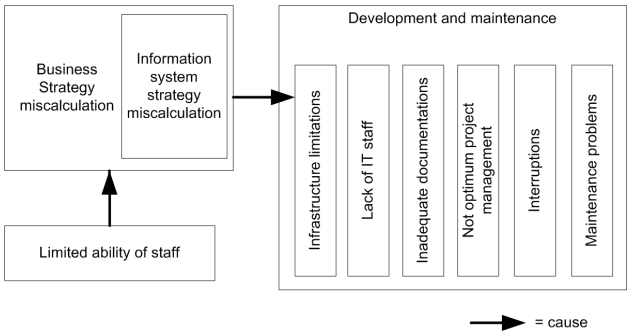


Figure 1. Problems at PCU Library IS Development Process

As shown in Figure 1 there are several problems faced by PCU Library IS development process such as business strategy miscalculation including IS strategy miscalculation. There was no clear analysis of strategic needs and information system environment. This caused miscalculations of PCU library business strategies [1].

These problems have caused several practical aspect problems at IS development and maintenance process. There were duplication in development efforts although the library had to cope with lack of IS staff, soon-will-be-obsolete technology platform, delays in finishing some of IS projects, and also not optimum quality of the developed applications. To overcome these problems there is a need to determine information system life cycle that fit with PCU library conditions.

2. PRINCIPLES OF ADAPTIVE INFORMATION SYSTEM LIFE CYCLE

Roger T. Burlton suggests ten principles of business process management [2]. Eighth principles states that every reform initiative should always be performed through a time-boxed and iterative approach. Seventh principle states that the process of change must be done from the outside in, and the sixth principle states that renewal initiatives should inspire thoughts that can be shared among people (shared insights).

In the eighth principle, Roger T. Burlton stated that any improvements will results in more errors before the organization can determine what are the right processes or paths. This means that any changes must be implemented first in order to know what the errors would be. Therefore it is necessary to set a short period of execution and then the evaluation is performed. This brief period is referred to as time-boxed execution and evaluation. Period of execution should not be too long that makes the efforts to improve will be too late. Execution and evaluation process will be performed over and over again in a short period of time that make the basis for improvement efforts. The main principle is incremental finding and execution. The knowledge gained from each iteration will be used to improve the next iterations.

Seventh principle states that improvement efforts should be based on particular levels. Without separation of improvement analysis in particular levels, the analysts will be stucked in "analysis paralysis". Improvement effort is initiated at the highest level, which then becomes more detailed at the next levels. The process of analysis begins by considering all stakeholders involved and pay attention to the interaction with the organization's stakeholders. Then the level of detail of process (analysis decomposition) is increased up to the level of activity.

The sixth principle states that the people within the organization who change. In any improvement step, organization must obtain knowledge, understanding, and innovative solutions required by the improvement process. But without the efforts to disseminate the knowledge, the knowledge will be owned only by certain people. There needs to be an effort to record and disseminate knowledges collected in each iteration of improvement so that

each iteration of improvement will be built on the knowledge gathered in previous iterations.

Therefore in order to create an adaptive business environment, an organization needs an adaptive information system, and also an adaptive development process. An adaptive development process needs to adhere several criterias:

1. Both processes of forming business strategy, and information systems development process should be divided into several levels of specific detail.
2. Each level will be carried out on separate iterations, and the previous iteration is the process of analysis (or development) of a higher level than the level at the current iteration (process decomposition).
3. Any knowledge from previous iteration is used for the purpose of achieving goals of current iteration (incremental finding and execution).
4. Each iteration is carried out in a short time (time-boxed) and evaluated at the end of each iteration (periodic evaluation).
5. Any knowledge gained in each iteration must be stored in order to be utilized in the next iterations (needs knowledge management system).

3. PROPOSED ADAPTIVE LIFE CYCLE

As the base of proposed adaptive IS life cycle (ISLC), this research uses 8 stages of system development life cycle as stated by Turban [3]. Eight stages of SDLC provides complete steps in planning, developing, implementation, auditing, and maintaining information system of the organization.

In order to be adaptive the eight stages of SDLC needs to consider changing business environment, so several stages have to be modified to include business process planning as seen in Figure 1.

Stage 2 and 3 of the SDLC are processes to establish information system design translated into seven separate stages in the proposed adaptive ISLC. In the adaptive ISLC, stage 2 to stage 8 are not only aiming to design information system, but also to form organization's business and information system strategy. The information systems strategy implements the business strategy.

Phase 4 of the eight stages of SDLC is translated into two separate stages in the adaptive ISLC, i.e. the stage 9 and 10. Stage 9 of the adaptive life cycle aims to establish a migration plan from current state of information systems into IS design that has been made at stage 7 and 8. Stage 8 will not produce actual information systems, so stage 9 is important to determine the sequence of the development of planned IS.

A complete stages of adaptive life cycle is shown in Figure 2.

3.1 Stage 1: Project Initiation

Project initiation is an effort to start the project. At this stage the head and members of the project are appointed. The scope of the project, as well as constraints of cost, time, and budget are established. At this stage it can also be determined the tools (software and hardware) that can be used by a team in doing their jobs.

3.2 Stage 2: Collecting Business Data and Current Information Systems

Adaptive information system life cycle begins with identifying the organization's business strategy, information systems strategy and then determines data, applications, and technology architectures necessary to implement the business strategy (stage 3-8). Once the entire set of architectures are established, information systems are developed (stages 9-10), implemented (stage 11), and maintained (stage 13).

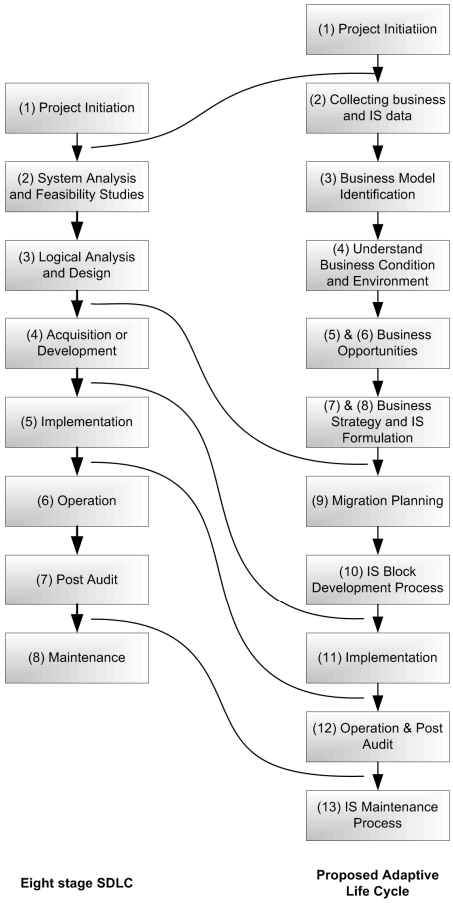


Figure 2. Information System Life Cycle

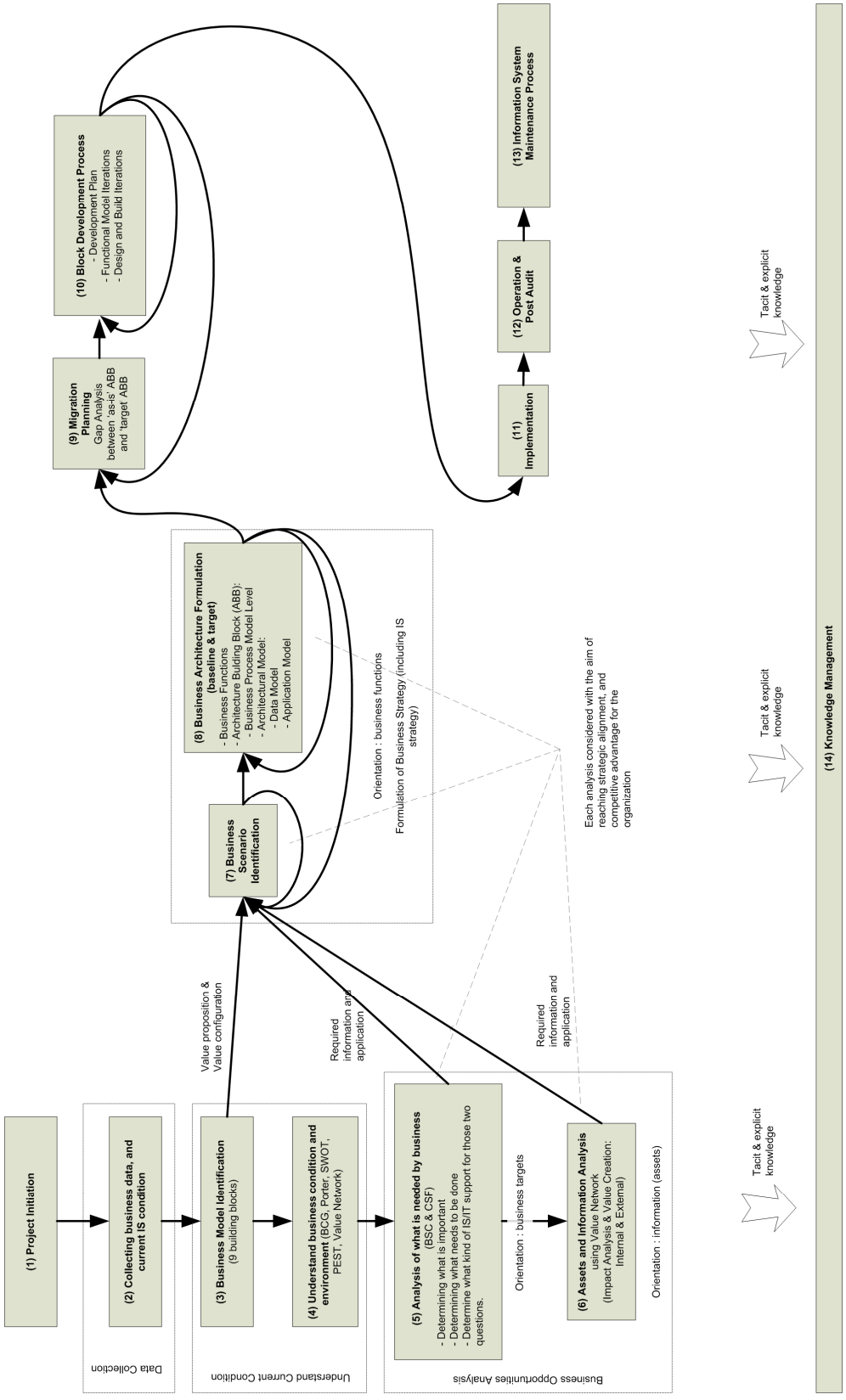


Figure 3. Complete Information System Life Cycle

Formulation of business strategy and information systems strategy are directed to achieve two targets suggested by John Ward [4]. Every analysis in this approach should be performed to achieve:

1. Strategic alignment. The alignment between business processes with information systems
2. Competitive advantage is to get an advantage against competitors using information systems.

To establish business strategy, necessary information about internal business environment and internal information system environment are needed. Information about internal business environment are:

1. Business strategy (not just business goals but the means to achieve business objectives): mission, vision, objectives, CSF (Critical Success Factor), business plan for every part of the organization.
2. Business processes, and data entities for each processes, Functional Decomposition Diagram, Process Flow Model, Entity Relationship Diagram, Data Flow Diagram, and Activity vs Entity Matrices. The main purpose of this step is to figure out the major functions, and minor functions performed by the organization. Business function is defined as a set of activities conducted in an effort to run a business [5].
3. Organizational environment, including organizational structure, assets and capabilities, and factors that are less visible (less tangible) such as knowledge, competencies, values, style, culture and relationships. All items can be molded into business principles that underlie all activities and strategies of the organization.

3.3 Stage 3: Identification of Business Model

After all information are gathered, the next step is to identify the business model of the organization. This step is to determine clearly what the business logic of the organization without being distracted from the short term strategy, or tactics of the organization. Analysts have identified nine pillars of the business model, including value proposition, target customer, distribution channel, relationship, value configuration, core competency, partner network, cost structure and revenue model. This step is very important because the pillars identified in this step will be used as an element of subsequent analysis steps.

3.4 Stage 4: Understand Business Condition and Environment

After the business model is identified, then analysis are performed to determine the condition of the organization, the external environment of business, and the organization position in the external business environment. Several tools can be used to perform the analysis such as Boston Consulting Group Matrix, Value Chain, Porter's Competitive Analysis, SWOT, and other tools.

Value Network is also used to describe all outstanding assets, both tangible and intangible assets, either in the form of products or services between internal entities within the organization, or with entities outside the organization. This stage uses Value Network instead of Value Chain [6] because:

1. Linear approach of inbound to outbound logistics is strongly influenced by processes in manufacturing industry. Whereas the conditions of modern business or assets of the flow of information can come from any place, and leads to any place too. Mechanistic and linear approach of the Value Chain is no longer adequate to describe the current production process.
2. Value Chain usually represents tangible values, but in today's world intangible value is also very important for the organization.
3. Value networks can be used on various types of organizations, both profit, and non-profit, manufacturing and service industries.

Value Network thus is more suitable for analyzing the flow of assets and information in various types of the organization.

3.5 Stage 5 & 6: Business Opportunities Analysis

After the analysis of business environment and internal information systems, stage 5 and 6 are performed to identify the opportunities that can be utilized by the organization. Stage 5 is oriented to business targets, and stage 6 is oriented to the exchanged information both within, and with parties outside the organization.

Analysis to find target-oriented business opportunity uses Balanced Scorecard (BSC), and Critical Success Factor (CSF). By using both tools, it must find:

1. Activities to be carried out by organizations that contribute directly to the achievement of business objectives. From these activities should then be determined the information, and applications needed to support these activities.
2. Activities to be undertaken by the organization to monitor and evaluate the activities from number 1. From this activity should also be found the information, and applications necessary to support, or allow such activity to take place (enabler).

The analysis of information-oriented opportunity uses Value Network diagram that was created in the second part of stage 4. By using the diagram from stage 4, two analysis are performed:

1. Value Impact Analysis. This analysis aims to find out problems, and opportunities gained from the information and assets (products, and services) received by organizations from outside entities. This analysis can help to find benefits gained by the organizations by making use of information not previously identified.
2. Value Creation Analysis. This analysis aims to determine the profits to be gained when organizations create value for other entities such as the loyalty of customers or suppliers.

3.6 Stage 7 & 8: Formulation of Business Strategy and IS Strategy

These stages use simplified The Open Group Architecture Framework (TOGAF) to form business strategy and information systems strategy. The formation process of business strategy and

information systems strategy begins by establishing business scenarios. Business Scenario analysis is an attempt to understand how the processes within the organization are progressed, who are involved, and requirements to be fulfilled in order to run the business scenario.

Business scenario may consists of:
1. Purpose of the scenario
2. Environmental and process models:
   a. Description of process
   b. Processes are mapped to the division of work
   c. The flow of information
3. Actors and responsibilities:
   a. Human actors, roles, and responsibilities
   b. Computer actors, roles, and responsibilities
4. Requirement of this scenario

In the PCU library, several examples of business scenarios are purchasing library materials from publishers, information dissemination process, and daily transactions of library materials. It is not required that the business scenarios reaches a certain level of detail in the initial iteration because it can be made more detail through the decomposition of business on the next iterations.

After initial business scenario has describes the entire organization process, although not in complete detail, then the process continues to the stage of formation of Business Architecture. Business Architecture is a model that describes the business structures which oriented in the business process, and information technology.

Business architecture consist of:
1. Organizational structure.
2. Business functions in the form of functional decomposition diagram.
3. Architecture Building Block (ABB), which consists of:
   a. Business Process Model.
   b. Architectural Model:
      i. Data block
      ii. Application block

The first step in the formation of business architecture is the identification of the organizational structure, and functional decomposition diagram. By using these two information, first iteration of business process model is created. First iteration of business process model only describes high level process of the organization and then can be used to create high level data block and application block to support the implementation of the business process model.

Next iterations will repeat the same steps for lower level of business scenarios. Acrhitecture building block (business process model, data block, and application block) will be revised in accordance with the details obtained from lower level business scenario. If ABB made from previous iteration is created for organizational level, next iterations will establish ABB for division level within the organization.

Formulation of business scenario and architecture building block iteration is repeated until architectural models (data, application) are adequate to be developed for the organization.

The process of formation of business architecture in a few iterations allow the use of creative or innovative principles that seek to answer the needs that have been identified at stage 5 and 6. At each iteration each person involved in the planning will seek to improve business processes with, or without, the involvement of information technology, and improve the architectural models that support these functions and business processes by leveraging the development of better technology.

3.7 Stage 9: Migration Planning

This stage compares current application portfolio (collected at stage 2), and target application portfolio made in stage 8. From target portfolio it should be identified which applications are new applications, which applications should undergo major changes, applications that have minor changes, applications that do not need to be changed, applications which are a combination of some old applications, and applications as a fraction of the old application. The list of applications that are no longer needed from current portfolio also need to be identified.

3.8 Stage 10: Architecture Block Development Process

In stage 10 the processes to develop each application block are started. The processes to develop application blocks use iterative development principles that each iteration will develop only small part of application requirements. This iterative development process allows less revision since previous iteration was performed in short time, and the evaluation process can be done immediately.

There are currently two major methods of application development that use iterative and incremental principle: the incremental model as used by the Unified Process (UP), and agile software development as used by Dynamic Software Development Model (DSDM), Scrum, and Extreme Programming (XP). These two models differ in the size of the application development iteration time. Each iteration of the incremental model like UP can take several months, while each iteration of the agile software development takes place in much shorter time, for example, during one week, two weeks, or one month. Because there is the need to be more adaptive, therefore it is decided use agile software development in the process of developing application blocks.

The three methods of agile software development are Dynamic System Development Method (DSDM) [7], Scrum [8], and Extreme Programming (XP) [9]. These three methods will be analyzed to determine which method is most suitable to be applied in the library of Petra Christian University, and whether it can be linked to other stages in the overall approach of this information system life cycle.

The analysis shows that:

1. All three types of method require a small number of programmers so that interaction and communication can go very intensively in the process of decision making at every stage of development. The number of programmers at Petra Christian University library amounted to no more than 5 people so PCU Library is very suitable for these methods.

2. Further analysis showed that the following conditions are not met, that is all of programmers should have qualification as

senior programmers. This qualification is very important in agile software development because every programmer should be trusted to make their own analysis and design after communication with the user. The less skilled programmers will produce a design application that is not easily understood by other programmers, and not scalable. Most of PCU library programmers are students so that the lack of experiences and insights into the development processes can reduce the quality of requirements analysis, and design of the application. DSDM provides the opportunity to reduce the problems that may arise because in DSDM the process of formation of functional requirements are placed at different phases. Scrum and XP performed all analysis and design steps on the same phase (they can even be done over and over again in a day if it is required). Given the separation of functional requirements iteration, and the design and build iteration, a senior programmer can do the prototyping process to record all functional and non functional requirements along with deciding the best methods to achieve these requirements. Junior programmers then can continue to build working prototypes according to the direction of senior programmer from the previous iteration.

3. Creating development documents is often not a major concern in the agile software development. There is no requirement analysis phase, analysis, and formal design, and programmers are advised to directly contact the user, analyze the request from the user, and immediately apply it to lines of code that makes the documents are often abandoned. Although the number of documents created remains much less than the incremental model, DSDM still provide a better approach than the two other agile methods. DSDM has business studies stage, and functional model iteration stage that formally analyze and produce documents as shown in Table 1 [7]. Scrum produced backlog and documents that are essentially derivatives of backlog, and XP generates user stories and the CRCs. These documents do not show all considerations that were taken in the process of analysis and design. To solve this problem every programmer can be asked to record all considerations taken in the analysis and design process and record it in any code written. Programmers can also use automatic documents generator tools that analyze source codes and interrelationships between each component in these lines of code.

Table 1. Documents Produced by DSDM

Phase	Documents Produced by DSDM
Business Study	1. Business Area Definition 2. Prioritised Requirements List 3. System Architecture Definition 4. Development Plan 5. Updated Risk Log
Functional Model Iteration	1. Functional Model (including Functional Prototypes) 2. Functional Model Review Records 3. Non-Functional Requirements List 4. Timebox Plans 5. Implementation Plan 6. Risk Log

4. All methods of agile software development are using short iterations ranging from one week to one month to produce a working prototype. Interactions with the user in each iteration are very intensive. Interactions are needed to determine functional and non functional requirements, and also to perform analysis, design, and testing. These interactions are so high that the criticism that arise for agile software development is that users who represent organizations in the team will have difficulty in doing the work for themselves. Non-IT related staffs at Petra Christian University library are relatively small in number so it will be difficult to ask one of the staff to be involved full time in the team. And also organization's representatives should be staffs that has high enough position to have insight and knowledge to make decisions, and also the authority to take decisions themselves without every time having to asking permission from the organization. DSDM gives flexibility that the iteration time length is longer than the two other methods, and the separation between functional model phase and design and build phase allows reduced interaction with the user.

5. All agile software development methods require that there are no interruption in the development iterations. Interruptions will disrupt the whole iterations' schedule, and for DSDM it can even disturb the whole project. Interruptions that may occur in the case of Petra Christian University library are data requests from the other PCU departments, and development of applications that previously have not even become a priority. To overcome this problem the library have to make a commitment that the development of new applications that previously have not been a priority to be carried out after a project is completed. And the library also have to create a tool that can be used by data processing division to provide data requested by other PCU departments without the need to involve the division of Software Development again.

By considering these five points of analysis, it was decided that Dynamic Software Development Method (DSDM) is more appropriate for the conditions in Petra Christian University library.

DSDM consists of five phases, coupled with a preliminary phase (pre-project), and the closing phase (post-project). The fifth phase is a feasibility study, business study, functional model iterations, design and build iterations, and implementation.

Feasibility study and business study phases have already been completed in stage 7 and 8 (formulation of business strategy and information systems strategy). Business area definition and system architecture definition that should be the products of the business study phase of DSDM has been produced in stage 8 as an architectural model. Each building block in the architectural model has described functions and business processes that must be met by the block, the parties involved (the divisions, and their responsibilities), the data architecture required to support the block (including operations on data - CRUD), and other applications that cooperate with the application block.

What has not been generated from stage 7 and 8 compared with DSDM phases is a development plan for each building block. Development plan is the result of the planning process that try to identify which prototypes needed to produce a working prototype. As with all methods of agile software development, the development process using prototypes that were developed in short iterations, and always get a review from the user. At the end of the development the working prototype gained acceptance from the user after going through the testing process. By using this method the user is always involved in every aspect of working prototype from requirements analysis, design, and testing.

In the early of stage 10 development plan is formulated. The objectives of formulation of development plan are:

1. Determine prototypes development priorities.
2. Identify the categories of prototypes necessary for the building block, and when the apropriate time to build them.
3. Decides the mechanism to determine when the prototype development activities should be stopped.
4. Determine individuals responsible in each development iteration.

Once the development plans are established then the iterations are conducted to produce functional prototypes and working prototypes. There are two types of iterations, i.e. functional model iterations, and design-and-build iterations. Functional model iterations are intended to produce documents and functional prototypes that defines the functional requirements, and non-functional requirements. Design-and-build iterations aims to implement functional prototypes to become working prototypes with high development standards specified in the coding principles. In the iterations all of non-functional requirements should also be met such as security issues, and performance.

DSDM also support the implementation of development projects management unlike Extreme Programming (XP). Extreme Programming has a development principle "what can be done tomorrow, do not do this now". It underlies every iteration of Extreme Programming so that user stories that have not been planned to be implemented will be stored and should not be considered in current development process, until finally the user story development schedule arrives. This results in no one can say when the entire project will be completed.

On the other hand DSDM must determine when the entire project must be completed through the outline plan, and when each iteration must be resolved through the development plan, and when the activities will be finished in each iteration through the timebox plan. The differences with traditional project management are the traditional project management establishes in exact detail all activities needs to be done, resources that must be provided, and the risks that must be addressed from the beginning of the project, even when the project has not been started yet. DSDM project management support incremental principle. At the beginning of the project sufficient detail are established so that the project can be started. If in the process more complete and detailed information are obtained (for example: protoype, and functional requirement) the project will be adjusted. If in an iteration not all requirements are expected to be completed, then the requirements that are not a must-have requirements (which are: should-have, could-have, and want-to-have-but-will-not-have-this-time) can be moved to the next iteration.

At the end of each design-and-build iteration the testing is always performed to check whether working prototypes has met the functional and non functional requirements. By using this approach then at the end of each iteration there are always new parts of product added without having to wait for the end of the project. Thus the user can trust the progress of the project better, and ensure that the end results will be in accordance with expectations.

After the development project for the building block (stage 10) is completed the life cycle return to stage 9. At stage 9 it needs to be determined what subsequent building block need to be developed according to development map that has been made previously. If all the building blocks necessary for implementation have been developed, then the life cycle can enter the stage 11 (implementation).

3.9 Stage 11: Implementation

Stage 11 converts old system into newly developed system. There are three ways that the organization can use to migrate from old system to new system. In the Parallel conversion old and new systems operate together during testing. When the new system considered to be stable and able to replace the old system, then the old system is stopped. In the Direct Cutover the old system is stopped and new system put into effect right away. This approach is most rapid, but also have the greatest risk. In Pilot Conversion the new system is implemented only in some divisions or branches. This approach is similar to the direct cutover on a particular branches, but for overall system it is like parallel conversion.

3.10 Stage 12: Operation and Post Audit

Stage 12 is the stage where the entire information system has been implemented, and begin to be used in daily operations by the organization. At this stage post audit is performed to evaluate the performance of the whole process of IS development process from stage to stage 1-11.

3.11 Stage 13: Maintenance

As can be seen in Figure 3 maintenance process uses chain of process that like a life cycle itself. It is advised to use maintenance process only to change bug fixing request, or minor enhancement, and not to change applications by requests that involves major changes in business process.

Each maintenance request will be classified to determine if it is a correction request (bug fixing) or enhancement request. Each request then analyzed to determine modification impact, alternative solutions, tests plan, and implementation plan. Maintenance team designs and develops changes needed by the application, tests the changes, and implements it to the organization.

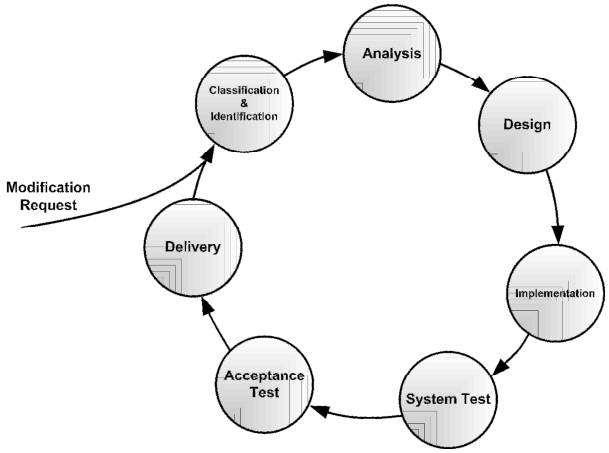


Figure 4. Maintenance Process

3.12 Knowledge Management System

Incremental finding and execution principle used in this approach needs that every knowledge gathered are stored to be used in the next iterations. To do this the organization needs knowledge expert to guide in knowledge capture process, decides formats suitable to store knowledge, create knowledge representations and knowledge base, tagging, and build knowledge taxonomy. This process will be addressed in future researchs.

4. CONCLUSION

This study proposes to expand 8 stages of SDLC into information system life cycle suitable to Petra Christian University Library. The information system life cycle uses principles such as incremental finding and execution, time boxed development, and knowledge management to record all knowledge gathered throughout the process. For PCU Library it is advised to use DSDM as a software development method with several modifications.

5. REFERENCES

[1] Wibowo, A. 2009. Faktor-faktor yang Mempengaruhi Kemampuan Sistem Informasi Beradaptasi Terhadap Perubahan Strategi Bisnis Organisasi: Perpustakaan Universitas Kristen Petra, *Design and Application of Technology*. Surabaya

[2] Burlton, R.T. 2001, Business Process Management : Profiting from Process, Sams Publishing

[3] Turban, E., McLean, E., Wetherbe, J. 2001, *Information Technology for Management*, 3rd Edition, John Wiley & Sons Inc.

[4] Ward, J., Peppard, J. 2002, *Strategic Planning for Information System*, 3rd, John Wiley and Sons, West Sussex, England

[5] Spewak. S. H. 1993. *Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology*, John Wiley and Sons

[6] Allee, V. 2002. A Value Network Approach for Modelling and Measuring Intangibles, *Transparent Enterprise*, Madrid.

[7] ---, *Product Overview*, DSDM Consortium, http://www.dsdm.org/version4/2/public/product_overview.html. Last accessed at April 3rd, 2012

[8] ---, Scrum for Team Systems - Guidance, http://www.scrumforteamsystem.co.uk/ProcessGuidance/v2. Last accessed at April 3rd, 2012

[9] ---, *Extreme Programming: A gentle Introduction*, http:/www.extremeprogramming.org. Last accessed at April 3rd, 2012

Secrets of Software Development and Project Management: Success or Failure

Deepak Murthy, Mohsen Beheshti
Computer Science Department
California State Univ. Dominguez Hills
1000 E. Victoria St. Carson, CA, USA
1-310-243-3398
mbeheshti @csudh.edu

Richard Alò
Computer & Mathematical Sciences
University of Houston Downtown
One Main Street, Houston, TX. USA
1-713-221-8207
alor@dt.uh.edu

Jianchao Han
Computer Science Department,
California State Univ. Dominguez Hills
1000 E. Victoria St. Carson, CA. USA
1-310-243-3398
jhan@csudh.edu

ABSTRACT
Majority of software developers, project managers, and management points their fingers at each other for their failures. Software projects fail because of many reasons. Every project is different from others. It could be either within the project itself, the organization, or outside in another company. Many researchers, practitioners, and textbooks have defined many methodologies, processes, and models that could have an impact on the success of completing a project effectively. In this paper, the key items for success and failures of software development and project management are investigated from three different perspectives: the perspective of software developments and developers, the perspective of software project managements, and the perspective of the project managers.

Keywords
Software lifecycle, software development, software process, project managements, project managers

1. INTRODUCTION
Failures happen for reasons. When a software project fails, most of the time management blames project managers and project managers blame software developers and vice-verse. In this paper we will often use entity in reference to software development and developers, project managers, or project management. No single project that we have heard/learned over the years in the software industry is successful or failed because of only one entity. The success of any project is interdependent on various entities. But if we ask a developer, manager, or management they always blame other entities why so and why not everyone takes responsibility for what has happened.

If the project doesn't work, look for the part that we didn't think was important [15]. Consider some of the reasons why projects have failed from the perspective of the entity software development and developers [12]: scope defined was wrong, technology used was wrong, methods used was wrong, team selected was wrong, vendor chosen was wrong, environmental disasters occurred at wrong time, team bad politics with management, requirements documented was wrong, poor team communication within, with product, with testing with management. There could be more and some specific to the project. How to overcome these problems and tips to make project successful will be analyzed and presented in detail in this paper.

The financial services authority has blamed "poor decisions" made by Royal Bank of Scotland's management and failings in the way the regulator itself oversaw the bank for the lender's near collapse [18]. Managements most of the times finger point other entities about the failures of projects. Some of the common management issues are - poor interpersonal and communication skills, ineffective team leaders, resistance to change, disappointing results, and inability to see beyond their functional silo. These problems are further discussed from the perspective of the project management in this paper.

If one asks for success and prepares for failure, he will get the situation he has prepared for. [13] Project manager didn't understand technology. The manager really didn't get the consequence of the decisions s/he was making, if only someone technical was in charge; or s/he micro-managed us the whole time. [14] Many reviews of failed projects have revealed many project managers were not good manager – but we've yet seen a project fail because the manager wasn't technical enough or spent too much time with the project. Some of the other reasons are - project manager is new to the company business or he might be doing project management for the first time, personal problems of project manager might lead to project failure, stakeholders giving wrong assurance, poor planning and estimation from project manager, treat project team badly and impact losing his team due to attrition, don't have good communication, don't have the ability to assess people and their capabilities. Such kinds of problems and more details shall be investigated from the perspective of project managers in this paper.

The rest of this paper is organized as follows. Section 2 describes the common problems why software projects fail and provides the practical tips for succeeding the projects from the perspective of software development and developers. The similar common problems and tips for improving the failures of software projects are investigated from the perspectives of project management and project manager in Section 3 and Section 4, respectively. Section 5 is the concluding remark.

2. SOFTWARE DEVELOPMENT

2.1 Why Software Development Fails
Developers are the backbone of the software. If the software is working perfectly with limited bugs, then its credit always goes to the developing team. Software often fails because of the developmental issues. None of developers is perfect when coming to software development. Project failures happen due to many

reasons. Some of the common problems that lead to projects failure or delay are summarized as follows from the perspective of software development and developers [2, 5, 8, 9, 11, 21, 22]:

1. Scope creep: When goals exceed the ability to deliver timely results project will fail for sure. Scope creep refers to uncontrolled and unexpected changes in user expectations and requirements as a project progresses. Some of the common mistakes that developers make are: overrunning initial cost estimations about developing the modules, over- or under-estimating project schedule and miscalculating work to personnel ratio.

2. Poor or non-functional architectural design: Any environment usually develops, and according to this development many issues may change such as strategies aligned to this environment objective, requirement, etc. It is clear and understood that what we are using today may be useless tomorrow. This concept should be considered when building any project. If the project architecture is inflexible for updates, the project may die by time because of daily changes and rapid developments.

3. Unrealistic expectations: One of the most common causes of friction between developers and their customers or managers is unrealistic expectations. Often customers simply start with unrealistic expectations since they are unaware of how much can be done in a project day. Sometimes project managers or developers inadvertently invite trouble by getting project approval based on optimistic estimates that don't take into account the reality of the "cone of uncertainty" caused by unknown requirements.

4. Lack of communication and collaboration: Developers tend to work without any interaction with other co-developers, which leads to miscommunication of building the wrong modules and, in turn, creates the delay with the project. Also many times developers have the ego and strange attitude that they show to other co-workers or managers which leads to project failure.

5. Resource dependencies and restrictions: In a large project where many developers are working, if a developer didn't develop a module in time, there would be dependencies on the other modules that need to be developed or integrated with other co-workers' modules. This might be due to various reasons, such as that developer didn't have prior experience in the technology or didn't foresee the issues. Sometimes developers also restrict themselves looking for help from other co-developers, which leads to project failures or delays of the projects.

6. Lack of visibility of who owns what and whether the development is on target: In a large project, often not any bit and piece of information about the project is known to all developers. It is hard to keep track of who owns the modules, if the project is not under central repositories. In such a situation, projects that have contractors/consultants or offshore teams often delay or fail to execute in time.

7. Failure due to integration (third party software or vendors) and inadequate testing environment: Developers often develop modules and merge their modules with vendors' software modules. These project often fail if they do not fully understand what the vendors' modules imply and not foresee what will happen during the merge. Developers have to rebuild entire system when this happens.

8. Failure due to financial system as it could not incorporate the developmental changes: Financial system modules often keep tending to change and developers need to be aware of these kinds and have to plan their module creation accordingly. Otherwise, the project tends to fail.

9. Failure due to management change: Developers won't be able to function when new managements take over the existing ones. New managements usually implement new rules and regulation that might affect the developers. For example, change the tradition project management to agile management such as Scrum. The project could tend to failure path if the change was resisted by the developers.

10. Failure due to attrition: Low salaries, poor facility, poor management, and project satisfaction lead to attrition and thus incomplete state of projects, which easily causes the failure or delay of projects.

2.2 Tips for Success of Software Development and Developers

Developers need to understand that there is no one solution to all projects, every manager is different, every developer they work with is different, every module they integrate is different, every tester is different, and every project needs to be treated as brand new project [1, 2, 12]. For successful development of the project, top ten common solutions are identified as success tips and listed as below:

1. Define clear requirements. Product team needs to provide the right documentation. Developers need to work closely with product team and provide the estimates to their managers about the changes to the new requirement. Always developers need to forecast the changes coming and need to think like a product person and need to suggest the changes and implementation cost to their managers.

2. Architectural design gets signed off by all stake holders. Before developing the modules, developers need to make sure if the underlying architecture for their modules is correct and need to discuss with experts in that domain. Often experts of the domains work as consultants and the developers can utilize them to perfect the architecture and think future of the architecture.

3. Define clear expectations. Every developer is different, so clear expectations should be defined for who does what and what modules should be delivered by individuals and when. As a freshman developer might not be able to deliver on time, the risk should be calculated during this phase. Developers need to understand the co-workers' modules and need to set the expectation of the co-workers to avoid developing incorrect modules.

4. Communication plays an important role. Often development is done remotely in offshore some times, or in different time zones. Developers, if working in different time zones, need to have effective communication so that the project information is up to date and all team members need to collaborate effectively and synchronize their work accordingly.

5. Resources should be aware of the technologies that are going to be used for the project. Developer training should be provided for new technologies to be used.

6. Developers need to develop milestones for the modules that are assigned to them individually and as a team. Instead of depending on the project manager, it always helps to keep track of what is being done and what needs to be completed during the process of developing a project.

7. Third party software and vendors should be integrated at the beginning of the project development. Adequate simulated testing environments should be available for testing.

8. New developmental changes need to be incorporated in financial systems at the beginning of the development or before the changes are applied.

9. Minimize or avoid management change during the project development cycle. If the change is inevitable, it should be tested on a subsystem for its effectiveness before implemented on the entire project.

10. Technical information is shared among all developers and all the technical work is clearly documented so that new resources or additional resources don't make any mistakes.

3. SOFTWARE PROJECT MANAGENENT

3.1 Why Software Projects Fail

Public decision makers may respond rather late to an emerging crisis. Management attempts to pin the blame for failure on external factors. The recent research reveals that more often than not the software project failure is the result of bad project decisions by leadership. Management fails to take right action at the right time and often projects fail due to this reason. The common problems/failures of software projects are summarized as below from the perspective of project management [6, 7, 10, 17, 19, 20, 24, 25, 26].

1. Poor communication: Management's poor/untimely communication causes failure of the project team reputation.

2. Inadequate leadership skills: Not having the right management head of the project leads to the project downfall.

3. Resistant to change: Management expresses frustration at the suggestion of change. Management's attitude and behavior signal skepticism. Management's attitude doesn't change even when the concerns have been addressed. Management is preoccupied with reminiscing about "what was" versus "what will be." Management continues to do things the same old way yet expects new results. Management is uncomfortable with ambiguity and isn't open to discovering better ways. Employees complain about mixed messages from leadership.

4. Poor vision: A poor vision allows members of an organization to disunite with individual goals, which leads to failure of the project.

5. Failure to assess direct and indirect competition: Often competitors watch the company progress and steal the ideas and release new products to customers before we release the product. Failure to assess this competition will lead to failure of the project.

6. Change in management at wrong time: Change in upper management directly impacts everyone in the company. There could be attritions, and many employees may not like the leadership of the new management and work deteriorates.

7. Management uses a wrong capital-to-labor proportion: Management doesn't minimize firm costs and maximize revenues. The capital should be used in proportion to the project spending and on the labor. Incorrect assumptions and extraneous spending lead to failure of the project.

8. Management operates inefficiently and uses technology ineffectively: Using poor technology for the products leads to failure of products.

9. Management's personal issues leads the project failure: Personal issues with the management and the investors can impact the failure of the project.

10. Shortages and lower customer satisfaction. Management doesn't fully take into account the customer's satisfaction while addressing the customer's functional requirements..

3.2 Tips for Success of Software Projects

Managing a software project is similar to operating a business. n the current economic situation it is challenging to keep the company business with profits. Listed below are some of the tips for the success of the management team to improve the company's business [1, 2, 4], which can be adapted to the management of software projects..

1. Empowering leadership: The first factor for business success is empowering leadership. This type of business management style has also been called transformational leadership. Transformational leadership is the type of motivational style that draws others in and inspires them to achieve something greater than themselves. However, the employees and staff members do not merely do the work; but they also become better people in the process.

2. Well-defined vision: The second factor to business success is a well-defined vision. A corporate vision is a scripted understanding of what a company wants to do and how they want to accomplish it. A well-defined vision allows members of a project team to unite for a common cause where singular aim and all energy focus in one direction.

3. Relevant knowledge of the business market: The third factor for business success is relevant knowledge of the business market. In order to do anything well, a person or company must do their homework to gain a deep understanding about the factors that are essential for success. These days as the World Wide Web continues to expand, there is no excuse for a would-be entrepreneur to lack knowledge of whatever business they feel led to pursue. Sadly, many businesses are dead out of the gate because they do not take the time to gain a proper perspective on the industry.

4. Detailed business plan: Another factor for business success is the formulation of a comprehensive business plan. Knowing about an industry and sketching out a vision is only the beginning of a successful enterprise. The next step is to take what we know and what we want to accomplish and write a detailed strategy for how to make it happen. A business plan covers all the related factors that are essential for a winning enterprise including

vision, description of the market, projected financials, employee relationships and customer relations management (CRM).

5. Assessment of the direct and indirect competition: When getting ready to implement a new project, another important factor for success or failure is the nature of the direct and indirect competition for the same product or service. For instance, a person or group wishing to open an online outlet for used furniture should take at least a few days to research how many other entities are trying to do the same thing. When doing the research the group should ask: who are the competitors? What products and services do they offer? What is their pricing structure? What kind of shipping do they offer? Gaining a firm grasp of the competition can definitely make the difference between staying alive long-term or filing for bankruptcy.

6. Availability of financing: The sixth factor important to the success or failure of a business is available financing. The current economic crisis in America has made venture capital difficult to find. Of course, if a company can manage to avoid using credit altogether then this is not a problem. Still, most new businesses need some kind of seed money to get them up to speed and thus the ability to secure working capital is critical to keeping the doors open.

7. Solid customer relations management: The seventh reason for project success or failure is how it relates to the customer. This seems like a no brainer, but the better a project handles the client, the more apt it will be to stay in business. Enterprises that take time to think out common and uncommon situations before they encounter them will be more likely to keep customers coming back. Those groups that merely define their customer relationships on the fly or in the heat of the moment are doomed to fail.

8. Proper timing: The eighth reason for project success or failure is timing. In 1998, when the latest housing boom began it was probably a good time to enter into the home mortgage industry; in 2006, when the housing bubble began to burst it was probably a poor time to set up a new mortgage outfit. Part of learning about an industry is getting a good feel for its business cycle; although trying to time the market can lead to indecision.

9. Well-devised decision making system: Decision-making is at the heart of any business and the best organizations outline a step of procedures involved in the decision-making process. Those entities that tend to practice participative leadership allow representatives from all departments to be involved in the process and seem to gain stronger employee buy in. Most poorly led organizations do not encourage participation and often lack a well-defined procedure for making decisions. One solid decision-making scheme is the nine-step problem-solving model:

- Describe the situation in detail
- Frame the "right" problem
- Describe the end-state goals from a broad perspective of values
- Identify the alternatives
- Evaluate the alternatives
- Identify and assess the risks
- Make the decision
- Develop and implement the solution
- Evaluate the results

10. Government regulatory measures: The tenth reason for business success or failure is how much the owners of an enterprise have a good grasp on the rules and regulations governing their sector of the economy; this includes having a clear understanding of the tax structure. Many would-be entrepreneurs charge into a good idea not knowing what restrictions apply to the execution of the idea. Not knowing the extent to government interference in an industry can mean the difference between success and failure of a business.

4. SOFTWARE PROJECT MANAGERS

4.1 Why Software Project Managers Fail

The top ten reasons why software project fails will be investigated from the perspective of project managers in this section. Individually, a project manager might be able to manage around any single cause, but taken collectively, a manager would have a colossal project management failure. In fact, any two of these reasons could seriously impact any project [3, 11, 14].

1. Bad communications: Project management is all about good communications. A manager has to keep everyone informed about changes, assumptions, requirements, standards, budgets, costs, and the schedule. Developing a good visual management system is a tremendous benefit to any project manager. Developing a good communication system will reduce the risk of project failure.

2. Poor schedule or resource management: Managing a project is really about managing the schedule, but a schedule is really a collection of resources that are being managed on a schedule. Mismanaging the resource and schedule will increase the chances of a project failing.

3. Weak requirements definitions (leading to inadequate planning): If we don't know where we are going then how do we know when we get there? A good project manager must know what the target is. The project requirements are the target. One way of defining requirements is to describe what the end result looks like in measureable or objective terms. Instead of requiring software to be friendly or easy to use, how about saying that it has to be simple enough that a 12 year old child can use it. We can argue about "easy to use" software but at least we now have an objective measure for the testing.

4. Inadequate planning, assumptions, risks, or resources: If we are planning a project, then we should be familiar with Murphy's Law [23], If anything can go wrong it will. Projects are frequently impacted by risks, assumptions about resource usage, or plain old surprises. The traditional solution is to add safety time or buffers to tasks to allow for schedule slippage and unplanned events. There are a lot of reasons why this fails too. We would suggest implementing the critical chain management to take more control over these buffers and not let them slip away.

5. Use of new or unproven technologies / methods: It can be so tempting to use the latest technology, or new method, but unless we are trained on the newest thing, all we are doing is introducing new risks. Projects have enough risks already and there is no need to introduce new risks from unproven technologies and/or methods.

6. Ineffective (or nonexistent) quality controls: Quality control is a simple system of checks and balances to ensure we are delivering what the customer asked for in the first place. What kind of quality control are we using for our project management? Are we tracking every project management nonconformance? Charting to find trends? Taking action when the trend is outside the norm? Most projects have a system for correcting problems but this is not the same as a system of corrective action. Project failures result from ineffective quality controls.

7. Managing multiple projects at once or multitasking resources: It seems so obvious to recapture downtime and juggle a few extra tasks in the meantime. But multitasking introduces complexity and schedule risk, which can impact all projects involved. Less is more. Focus is power. We can actually increase our productivity by focusing on fewer projects.

8. Supply chain failures: Sometimes we have to contract out the work. But managing contractors has inherent risks too. Contractors may not have the skill level required for the task. Contractors may have multi-tasks in order to stay busy but the multitasking introduces complexity and schedule risk. It sounds so easy to contract out some of the work to the supply chain but it actually adds to complexity unless we have clearly defined requirements agreed to first.

9. Scope creep or poor impact analysis: This happens on practically every project, doesn't it? We start with a clear concept or at least we thought it was clear when we started. Then one thing leads to another and before we know it we are involved in a different project. That's scope creep. If we are doing "A" then we must do "B" and if we are doing "B" then we have to also do "C". Clarifying the real requirements and performing a good impact analysis are two methods to solve this problem. Otherwise, scope creep will impact the schedule, the budget, and the resources.

10. Lack of qualified resources: We will just have to make do. If we hear this then we know we are questioning the resources. Assigning the wrong task to the wrong person will be impacting the project. This usually happens because we lack enough experienced resources.

4.2 Tips for Success of Software Project Managers

Tips listed below for success of software project managers are general and are not specific to any projects. Each project is different, and project managers need to define the process for each project [14]. Managers also need to constantly be aware of the problems listed above and need to have a clear understanding of their environment. The following tips can be used for managers to improve their process of managing the software projects [1, 2, 14, 16, 22].

1. Project managers always need to emphasize for clear and complete requirements from product teams for building right software. A technically sound project manager often makes the right decision about sending the requirements back to the product teams to get the right user requirements. If the manager is not technically sound, s/he needs to take the team members who are experienced in understanding the requirements to help the manager estimate the project accurately.

2. Project managers should give more attention to minutest detail of planning and in preparing the planning document. Planning document should be made as a living breathing document until the completion of the project cycle.

3. Project managers always need to improve the current process with new tools and techniques that are used commonly in the industries. Some times, trying new things also delays the project, instead they need to get trained on new tools for management and often need to bring in specialized management tools and calibrate them to their experience.

4. By introducing modern metrics and measures in place to supplement enhanced controls will improve project managers' capabilities and create an atmosphere of trust with senior management.

5. Managers need to put effort today into developing the infrastructure that will be necessary to generate effective products tomorrow.

6. Managers need to focus on explaining why or why not the software is different than other technical disciplines.

7. Managers need to always put the right product into the market. No one prospers when defective products are released to market. Customers and users demand that we do a better job when it comes to quality.

8. Project managers need to interact with the executive committee and always stay on top of getting approvals from them to satisfy the developers and the product teams.

9. Project managers need to be the first persons to try new process or need to create a small research team that successfully shows the results and stands as an example, before implementing them on the entire teams.

10. Project managers need to plan estimation and disaster recovery plans or risk analysis plans for their projects before starting the projects. This helps managers go back to management teams in buying times or crunching existing times and making developers work intelligently to complete the tasks.

5. CONCLUDING REMARK

In this paper we have discussed, investigated, and summarized the common problems and success tips to improve software project from three different perspectives of software development/developers, software project management, and software project managers, respectively. These failure problems and success tips are general-purposes. For certain specific cases, the tips would not apply. In our future research we would like to take a case study and show how to effectively apply the tips and avoid the problems. Also we would like to expand the problems and tips for testing quality control and also software ethics.

6. ACKNOWLEDGMENTS

This paper is based on work supported by the Department of Homeland Security under the auspices of the Southeastern Regional Visualization and Analytics Center, Cooperative Agreement No. 2009-ST-061-CI0001, and National Science Foundation (NSF) through grant CNS-05040538. Any opinions,

findings, and conclusions or recommendations expressed in the paper are those of the authors only

7. REFERENCES

[1] Agarwal, N. & Rathod, U. (2006). Defining 'success' for software projects: An exploratory revelation. International journal of project management, 24, 358-370.

[2] Donald J Reifer, Barry Bohem. (2006) Software Management, 7th ed., Wiley-IEEE Computer Society Press.

[3] Chris Anderson. Top 10 Causes of Project Management Failures from http://www.bizmanualz.com

[4] Ecoggins, 10 Reasons for Business Success or Failure from http://ecoggins.hubpages.com.

[5] Damian, D. & Moitra, D. (2006). Global software development: how far have we come? IEEE Software, September/October 2006.

[6] Holt, R., 2004. Risk management: the talking cure. Organization 11 (2),251-270.

[7] Karel de Bakker, Albert Boonstra, Hans Wortmann. (2009). Does risk management contribute to IT project success? A meta-analysis of empirical evidence. ijproman

[8] Kraut, R.E. & Streeter, L. (1995). Coordination in software development. Communications of the ACM, 38(3).

[9] Kuni, R. & Bhushan, N. (2006). IT application assessment model for global software development, IEEE International conference on global software engineering, 2006.

[10] Kwak, Y.H., Stoddard, J. (2004). Project risk management; lessons learned from software development environment. Technovation 24, 915-920.

[11] Layman, L., Williams, L., Damian, D., & Bures, H. (2006). Essential communication practices for extreme programming in a software development team. Information and software technology, 48, 781-794.

[12] Linberg, K.R. (1999). Software developer perceptions about software project failure: a case study. The journal of systems and software, 49, 177-192.

[13] Procaccino, J., Verner, J., Overmyer, S., Darter, M. (2002). Case study: factors for early prediction of software development success. Information and Software Technology 44, 53-62.

[14] Procaccino, J.D., Verner, J.M. (2006). Software project managers and project success: an exploratory study. The Journal of Systems and Software 79, 1541-1551.

[15] Project Management Institute (2004). A Guide to the Project Management Body of Knowledge, Project Management Institute, Newtown Square, PA.

[16] Ravichandran, R. & Ahmed, N.U. (1993). Offshore systems development. Information & management, 24, 33-40.

[17] Raz, T., Shenhar, A.J., Dvir, D. (2002). Risk management, project success, and technological uncertainty. R&D Management 32 (2), 101-109.

[18] RBS report: 'poor decisions' by management and FSA blamed for failure. Dec 15 2012 from http://www.telegraph.co.uk

[19] Ropponen, J., Lyytinen, K. (1997). Can software risk management improve system development; an exploratory study. European Journal of Information Systems 6, 41-50.

[20] Ropponen, J., Lyytinen, K. (2000). Components of software development risk: how to address them? A project manager survey. IEEE Trans. on Software Engineering 26 (2), 98-112.

[21] Sabherwal, R. (2003). The evolution of coordination in outsourced software development projects: a comparison of customer and vendor perspectives. Information and organization, 13, pp. 153-202.

[22] Sole, D. & Edmondson, A. (2002). Situated knowledge and learning in dispersed teams. British journal of management, 13, S17-S34.

[23] Sommerville, I. (2011). Software Engineering, 9th ed., Addison-Wesley.

[24] Wallace, L., Keil, M., Rai, A. (2004). Understanding software project risk: a cluster analysis. Information and Management 42, 115-125.

[25] Wallace, L., Keil, M., Rai, A. (2004). How software project risk affects project performance. an investigation of the dimensions of risk and an exploratory model. Decision Sciences 35 (2), 289-321.

[26] Weick, K., Sutcliffe, K. (2007). Managing the Unexpected, 2nd ed. Wiley, New York.

[27] Zafiropoulos, I., Metaxiotis, K., Askounis, D. (2005). Dynamic risk management system for the modeling, optimal adaptation and implementation of an ERP system. Information Management and Computer Security 13 (3), 212-223.

Multi-faults diagnosis of rotor-bearing systems using Hilbert-Huang spectrum and FFT

Weidong Jiao
Department of Mechanical Engineering
Jiaxing University, Jiaxing, China
118 Jiahang Road, 314001.
TEL: 086-573- 83644860.
Fax: 086-573-83644860.
jiaowd1970@mail.zjxu.edu.cn

Suxiang Qian
Department of Mechanical Engineering
Jiaxing University, Jiaxing, China
118 Jiahang Road, 314001.
TEL: 086-573- 83644860.
Fax: 086-573-83644860.

Yongping Chang
Department of Mechanical Engineering
Jiaxing University, Jiaxing, China
118 Jiahang Road, 314001.
TEL: 086-573- 83644860.
Fax: 086-573-83644860

Shixi Yang
Department of Mechanical Engineering and Automation
Zhejiang University, Hangzhou, China
Zheda Road 38, 310027.
TEL: 086-571-87951924.
Fax: 086-571-87951145.
yangsx@zju.edu.cn

Gongbiao Yan
Department of Mechanical Engineering and Automation
Zhejiang University, Hangzhou, China
Zheda Road 38, 310027.
TEL: 086-571-87951924.
Fax: 086-571-87951145.

ABSTRACT
Unbalance is the most common fault in rotor system. Fatigue crack and rotor-stator rub are also the two important rotor faults. Vibration behavior of a faulted rotor often shows strong nonlinear and transient characteristic, especially when response-dependent nonlinear breathing crack model is considered in a cracked rotor, and/or there are other faults such as rotor-stator rub in the system. The article is aimed to study novelty feature extraction from time-frequency distributions of the vibration response of the faulted rotor with multi-faults such as unbalance, fatigue crack and/or rotor-stator rub, using Hilbert-Huang spectrum analysis. Equations of motion of the faulted Jeffcott rotor with unbalance, transverse crack and rotor-stator rub are presented. By solving the motion equations, steady-state vibration response is obtained in presence of multi-faults. Hilbert-Huang spectrum analysis based on Hilbert-Huang transformation is used to find the distinctive feature of the rotor system when these rotor faults exist independently or together. Based on the study some diagnostic recommendations are derived.

Keywords
Multi-faults; Stiffness coefficients; Empirical mode decomposition; Instantaneous frequency; Time-frequency representation; Fault diagnosis.

1. INTRODUCTION
In essence, fault diagnosis can be looked on as a typical task of pattern recognition, in which feature extraction is inevitably involved and plays a crucial role [1]. Therefore, it is important for fault diagnosis of rotor system to extract representative features from time-domain vibration response belonging to different fault patterns (including normal condition, different faults and their different combinations). With rotors become more flexible and operate under tight clearances and harsh environment in the present day, fault diagnosis of rotor system becomes more complicated, and thus difficult, especially when there exist more faults than one in the system.

A rotor system can always have some amount of unbalance, misalignment in the driveline, temporary bow, etc. However, these can be within the permissible limits. Furthermore, these initial drawbacks or faults may excite one or more faults in the system under such circumstances. For example, rotor might develop contact with the stationary parts under tighter clearances or rotor could develop fatigue crack under severe thermal and mechanical stresses [2]. These two important rotor faults, i.e. fatigue crack and rotor-stator rub, could lead to catastrophic failure, if remain undetected. Many researches have been made on fault diagnosis of crack and rotor-stator rub. Zhou *et al*. [3] studied the nonlinear dynamical behavior of a horizontal Jeffcott rotor system with a transverse crack. They used a new acceleration scheme to reduce shocks to the rotor and diagnose the existence of cracks. Sabnavis *et al*. [4] made extensive reviews of the literature on vibration characteristics of cracked rotor shaft. Muszynska [5, 6] made excellent researches and reviews on rotor-stator contacting phenomenon and rub-related diagnostics. She found contacting interaction between rotor and stator is manifested by one or more of the physical phenomena, i.e. friction, impacting and stiffening, depending upon system

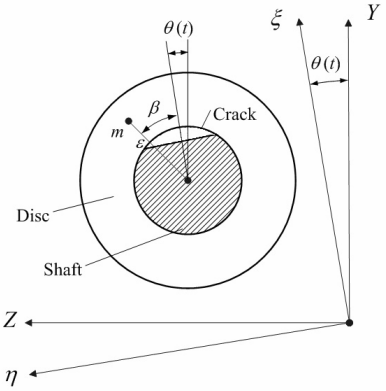
parameters and operating conditions, different types of rubbing conditions prevail. Feng and Zhang [7] studied vibration response of a rotor rubbing caused by an initial perturbation. They discussed influence of various parameters on the vibration phenomena of the rotor system. Chu and Lu [8] observed very rich form of periodic and chaotic vibrations in their experimental study. Based on the model-based diagnostic approaches, Bachschmid and Pennacchi [9], and Platz *et al*. [10] studied multiple faults identification of rotor system. Darpe [11] applied the signal-based approach for analysis on the vibration response of the rotor system with crack and asymmetry faults together. Recently, Patel and Darpe [2, 12] have numerically and experimentally investigated the rotor vibration characteristics of the unbalance, crack and rotor-stator rub faults, using both classical Jeffcott rotor model and Timoshenko beam element.

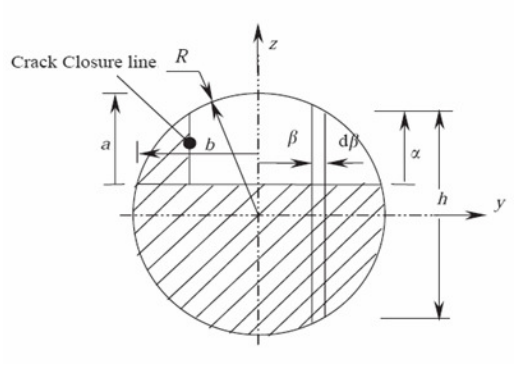
In this study, steady-state vibration response of a multi-faults (including unbalance, crack and rotor-stator rub) Jeffcott rotor supported on simple rigid bearings is investigated, considering the fact that rotors become more flexible in the present day. Our primary objective is to investigate nonlinear and nonstationary characteristics of the rotor vibration in presence of multi-faults. This study utilizes the crack model and the rotor-stator rub model presented in Ref. [12]. Since the rub fault is highly nonlinear and transient, Hilbert-Huang transformation (HHT) is used for the study along with the conventional fast Fourier transformation (FFT).

2. VIBRATION RESPONSE OF CRACKED ROTOR WITH UNBALANCE AND/OR ROTOR- STATOR RUB

A Jeffcott rotor is considered on rigid-bearing supports having a central disk of mass m on the shaft of length L and diameter D. Transverse surface crack of depth a, is assumed at the mid span of the rotor. Fig. 1(a) shows the coordinates (Y and Z are the fixed coordinates and ξ and η are the rotating coordinates) in the crack cross section.

(a)





(b)

Figure 1. (a) Coordinate system, and (b) cross-section at crack location [12]

The unbalance eccentricity ε is assumed to be at an angle β with the weak crack direction. $\theta(t)$ is the instantaneous angle of rotation and ω is the rotor speed. Rotor–stator interactions forces are considered at disc location. The equations of motion for the cracked Jeffcott rotor, can be written in fixed coordinate Y–Z system (Fig. 1(a)) as [12]

$$\begin{aligned} &m\ddot{Y}+c\dot{Y}+\frac{1}{2}\left\{k_\xi+k_\eta+\left(k_\xi-k_\eta\right)\cos(2\theta)-2k_{\xi\eta}\sin(2\theta)\right\}Y \\ &+\frac{1}{2}\left\{\left(k_\xi-k_\eta\right)\sin(2\theta)+2k_{\xi\eta}\cos(2\theta)\right\}Z=F_y+m\varepsilon\omega^2\cos(\theta+\beta)-mg, \\ &m\ddot{Z}+c\dot{Z}+\frac{1}{2}\left\{k_\xi+k_\eta+\left(k_\eta-k_\xi\right)\cos(2\theta)-2k_{\eta\xi}\sin(2\theta)\right\}Z \\ &+\frac{1}{2}\left\{\left(k_\xi-k_\eta\right)\sin(2\theta)+2k_{\xi\eta}\cos(2\theta)\right\}Y=F_z+m\varepsilon\omega^2\sin(\theta+\beta). \end{aligned} \tag{1}$$

The forces F_y and F_z are the nonlinear rub forces. k_y and k_z are direct and k_{yz} and k_{zy} are cross-stiffness coefficients of the cracked rotor, defined in the fixed coordinate system. These coefficients are calculated from the stiffness coefficients k_ξ, k_η, $k_{\xi\eta}$ and $k_{\eta\xi}$ defined in the rotational frame in a cross-section containing the transverse fatigue crack. In breathing crack model considered for this study, amount of open part of the crack continuously changes with shaft rotation, thereby accounting partial open/close state of the crack. Due to this partial opening and closing of the crack, cross-stiffness terms $k_{\xi\eta}$ and $k_{\eta\xi}$ appear in the equations of motion (Equation 1, as $k_{\xi\eta}$ is related to k_{yz}).

2.1 Calculation of stiffness coefficients k_ξ, $k_{\eta\xi}$, and $k_{\xi\eta}$ ($k_{\eta\xi}$)

The stiffness coefficients k_ξ, k_η and $k_{\xi\eta}$ ($k_{\eta\xi}$) is found out from the direct and cross-flexibility coefficients g_ξ, g_η, and $g_{\xi\eta}$ ($g_{\eta\xi}$) of the cracked shaft, using linear elastic fracture mechanics theory: Castingliano's theorem [12], as follows:

$$k_\xi=\frac{g_\eta}{g_\xi g_\eta-g_{\xi\eta}^2},\quad k_\eta=\frac{g_\xi}{g_\eta g_\xi-g_{\eta\xi}^2},\quad k_{\xi\eta}=k_{\eta\xi}=\frac{-g_{\xi\eta}}{g_\xi g_\eta-g_{\xi\eta}^2}. \tag{2}$$

The flexibility coefficients can be found using Castingliano's theorem using the strain energy of the cracked shaft. The strain energy of the rotor in presence of crack can be found by using the strain energies of the uncracked shaft and the additional strain energy due to the presence of crack. Total flexibility of the cracked shaft made up of two parts: one is the flexibility of uncracked shaft and second is the additional flexibility introduced by the crack. The flexibility introduced by crack, changes with the amount of the open part of crack. As the rotor rotates, the crack breathes. The amount of open part of the crack constantly changes, thereby changing the flexibility of the cracked rotor. The flexibilities are computed using [12]

$$g_{\xi} = \frac{L^3}{48EI} + \iint \frac{128L^2\alpha h^2}{\pi ED^8} F(\alpha/h)d\alpha d\beta$$
$$g_{\eta} = \frac{L^3}{48EI} + \iint \frac{512L^2\alpha\beta^2}{\pi ED^8} F'(\alpha/h)d\alpha d\beta$$
$$g_{\xi\eta} = g_{\eta\xi} = \iint \frac{256L^2\beta h^2}{\pi ED^8} \alpha F(\alpha/h)F'(\alpha/h)d\alpha d\beta \quad (3)$$

Where the term $L_3/48EI$ represents the flexibility of the uncracked shaft, and functions F and F' are given by Ref. [12].

The crack breathing condition has significant influence on the variation of stiffness coefficients, which is found using the sign of the overall stress intensity factor (SIF) [2]. Integration limits in expressions of Eq. (3) are from 0 to α for depth of crack and the limits for width are not specified (Fig. 1(b)). The limits of integration on width depend on the open part of the crack, which can be obtained from the sign of SIF along the crack edge. The negative sign of SIF indicates compressive stress and the closed crack condition and positive sign indicates tensile state of stress and open crack condition. Accordingly, depending on the instantaneous state of forces acting on the crack element, the SIF is influenced, which influences the stiffness coefficients. The detailed derivation of the flexibility coefficients of the cracked rotor can be found in Ref. [12].

2.2 Rotor-stator rub forces

When rub occurs, the impact force F_y and F_z can be written as

$$F_y = -F_N(y/e_r) - \psi_f F_T(z/e_r), \quad F_z = -F_N(z/e_r) + \psi_f F_T(y/e_r) \quad (4)$$

Where ψ_f is the function that decides the direction of frictional forces

$$\psi_f = \begin{cases} -1 & for \ \ \omega R + v_t > 0 \\ 0 & for \ \ \omega R + v_t = 0 \\ +1 & for \ \ \omega R + v_t < 0 \end{cases} \quad and \ \ v_t = \dot{Z}\left(\frac{Y}{e}\right) - \dot{Y}\left(\frac{Z}{e}\right) \quad (5)$$

Where R is the disc radius and v_t is the tangential velocity at disc location.

2.3 Hilbert-Huang transformation (HHT)

The HHT technique, consisting of empirical mode decomposition (EMD) and the well-known Hilbert transform (HT), was proposed by Norden *et al*. [13] in 1998. Using the EMD, one can decompose any complicated vibration signal into finite number of intrinsic mode functions (IMFs). Such decomposition is completely adaptive and the extracted IMFs are almost orthogonal. Then, the HT is applied to every IMF component to produce time-frequency-energy distribution of the vibration data together, i.e. Hilbert–Huang spectrum (HHS). Theoretically, HHT technique is adaptive and able to provide equal resolution at all frequencies and time instants, which makes use of HHT more meaningful for transient vibration signals such as rotor coast up vibration response. As a powerful tool for analysis of nonlinear and nonstationary signal, the HHT technique has successfully been used in signal analysis of faulted rotor system [14, 15].

After the EMD procedure, the original signal can be represented as,

$$y(t) = \sum_{j=1}^{n} c_i + r_n \quad (6)$$

After performing the Hilbert transform to each IMF component $c_i(t)$, the original signal can be expressed as the real part (RP) in the following form:

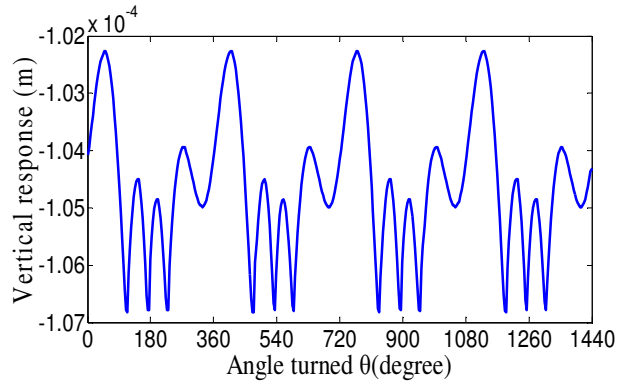
$$y(t) = RP\sum_{i=1}^{n} a_i(t)e^{j\phi_i(t)} = RP\sum_{i=1}^{n} a_i(t)e^{j\int \omega_i(t)dt} \quad (7)$$

Here the residue r_n is left out as it is either a monotonic function or a constant. Eq. (7) gives both amplitude and frequency of each component as functions of time. Thus, time–frequency–energy map of the original signal $y(t)$ can be obtained, which is known as Hilbert–Huang spectrum.

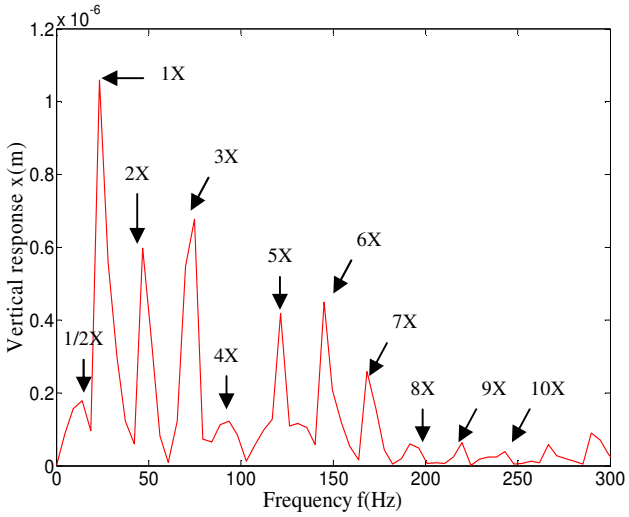
3. NUMERIC RESULTS AND DISCUSSIONS

Values of the system parameter in our numerical simulations are selected below, according to Ref [12]: Mass of the disk, m = 6 kg; Shaft diameter, $D = 2R$ = 0.025 m; Crack depth a/D = 0.15; Shaft length, L = 0.7 m; External damping, c = 182.56 Ns/m; Stator stiffness, k_s = $140\times10^{+6}$ N/m; Coefficient of friction, μ = 0.2; Unbalance eccentricity, ε = 1.0597×10^{-5} m. Bending natural frequency of the uncracked rotor, ω_{cr} = 48.42 Hz; Rotating frequency of rotor shaft, $\omega = 0.5\omega_{cr}$ = 24.21 Hz. This section discusses spectral characteristic of the unbalance vibration responses of the cracked rotor without rotor-stator rub and of the uncracked rotor with rotor-stator rub, using conventional FFT and HHS plots. Moreover, time-frequency spectral characteristic of the unbalance rotor response is identified, when rub and crack coexist.

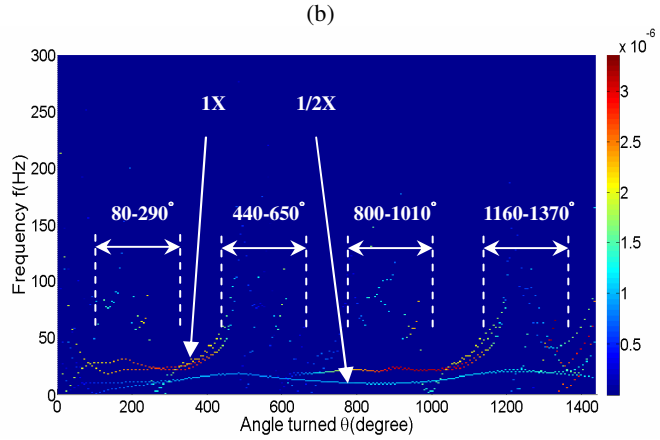
Unbalance vibration responses of the cracked rotor in presence of rotor-stator rub is now discussed. Fig. 2 shows the time-domain response in vertical direction, FFT spectral and HSS plots, for the cracked rotor with rotor-stator rub.



(a)



(b)



(c)

Figure 2. Vibration response, FFT spectral and HHS plots of cracked rotor (*a/D* = 0.15) with rub: (a) vertical vibration response; (b) FFT spectral plot, and (c) HHS plot.

From the vibration waveform and FFT plot (Fig. 2), it may be noted that the harmonics of 1X order (24.21 Hz) exists in the response. The FFT spectral and HHS plots show that the higher frequencies get excited as rub takes place. These frequencies disappear once the rotor loses the contact. In this case, it may be noted that higher level of vibrations, due to presence of crack, leads to more much interactions between rotor and stator, compared with the case of uncracked rotor with rotor-stator rub. The rub occurs for 80–290°, 440–650°, 800–1010°, and 1160–1370° of shaft rotation angular positions. The HHS plot correspondingly shows higher harmonics in the vertical vibration with fluctuation in the excitation frequency, which is a typical rub feature. The higher frequencies disappear when there is no rub, for example at 290–440°, 650–800° and 1010–1160° of shaft rotation.

It should be noted that the 2X, 3X, and 5X harmonics are mainly due to crack and harmonics at 4X, 6X, etc. arise due to rotor-stator rub [12]. These harmonics are clearly seen in FFT as well as HHS plots (Fig. 2 (b) and (c)). Presence of strong 2X harmonic frequency component at respective subharmonic resonance is typical of cracked rotor and backward whirling nature of vibrations at this harmonic is attributed to the rotor-stator rub phenomena. In case of rub in cracked rotor, 2X, 3X, and 5X frequency components are sufficiently strong in FFT spectra and HHS plots. Hence, from diagnostics point of view, it is important to monitor the nature of motion of 2X frequency as well as higher harmonics at corresponding subharmonic resonances.

4. CONCLUSIONS

In this paper, one multiple faults case, i.e. unbalance, crack, and rotor-stator rub, in a rotor system is considered. Numerical simulation is carried out and vibration signals are analyzed using the HHT based method for time-frequency representation. The study brings out the nonlinear and non-harmonic nature of the spectral components for the rotor faults. The nonlinear dynamic response due to rotor-stator interaction that appears in vibration waveform is also revealed by HHT, with complete information in the form of time, frequency and associated energy.

From the investigations, relatively new diagnostic recommendations are proposed. Since rub-impact is transient in nature, it gives broadband excitation and the spectrum of the rub signals shows numerous spectral lines. It has been shown that rub in cracked rotor excites higher frequency components with almost equally periodical mode and this is suggested as fault feature typical of rotor crack and rub. It is also important to observer the nature of the 2X vibration motion. It has been shown that when rub develops in the cracked rotor, the original strong 2X vibration motion of the cracked rotor (typical of crack fault) becomes relatively weak. The nature of higher harmonics at corresponding subharmonic resonances i.e. 3X, 4X and 5X etc. is also recommended to identify rotor-stator rub in a cracked rotor.

5. REFERENCES

[1] X. Min. Manual of machinery fault diagnosis. Xi'an Jiao tong University Press, Xi'an, China, 1998.

[2] T. H. Patel, and A. K. Darpe. Coupled bending-torsional vibration analysis of rotor with rub and crack. Journal of Sound and Vibration, 2009, 326: 740-752.

[3] T. Zhou, J. X. Xu, and Z. C. Sun. Dynamic Analysis and Diagnosis of a Cracked Rotor. Journal of Vibration and Acoustics, 2001, 123(4): 539-543.

[4] G. Sabnavis, R.G. Kirk, M. Kasarda, and D. Quinn. Cracked shaft detection and diagnostics: a literature review. The Shock and Vibration Digest, 2004, 36: 287-296.

[5] A. Muszynska. Rotor-to-Stationary Part Rubbing Contact in Rotating Machinery. Rotordynamics, CRC Press, Taylor & Francis, Boca Raton, 2005, pp.555-710.

[6] A. Muszynska. Rotordynamics. CRC Press, Boca Raton, FL, Taylor & Francis, London, 2005.

[7] Z. Feng, and X. Zhang. Rubbing phenomena in rotor–stator contact. Chaos, Solutions and Fractals, 2002, 14: 257-267.

[8] F. Chu, and W. Lu. Experimental observation of nonlinear vibrations in a rub-impact rotor system. Journal of Sound and Vibration, 2005, 283: 621-643.

[9] N. Bachschmid, and P. Pennacchi. Model based malfunction identification from bearing measurements. Proceedings of Seventh In Conference on Vibration in Rotating Machinery, Nottingham, 2000, pp. 571-580.

[10] R. Platz, R. Markert, and M. Seidler. Validation of online diagnostics of malfunctions in rotor systems. Proceedings of Seventh In Conference on Vibration in Rotating Machinery, Nottingham, 2000, pp. 581-590.

[11] A. K. Darpe. Dynamics of Cracked Rotor. PhD Thesis, Department of Mechanical Engineering, IIT Delhi, New Delhi, 2000.

[12] T. H. Patel, and A. K. Darpe. Vibration response of a cracked rotor in presence of rotor–stator rub. Journal of Sound and Vibration, 2008, 317: 841-865.

[13] E. H. Norden, S. Zheng, R. L. Steven, *et al*. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of Royal Society, London. A, 1998, 454: 903-995.

[14] D. Yu, J. Cheng, and Y. Yang. Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mechanical Systems and Signal Processing, 2005, 19: 259-270.

[15] K. Qi, Z. He, and Y. Zi. Cosine window-based boundary processing method for EMD and its application in rubbing fault diagnosis. Mechanical Systems and Signal Processing, 2007, 21: 2750-2760.

Printer on Garment Printing

Rahmadi Trimananda
Universitas Pelita Harapan
M.H. Thamrin Boulevard 00-00
Lippo Karawaci, Tangerang 15811, Indonesia
Phone: +62-21-546-0901 ext.2300
rahmadi.trimananda@staff.uph.edu

Arnold Aribowo
Universitas Pelita Harapan
M.H. Thamrin Boulevard 00-00
Lippo Karawaci, Tangerang 15811, Indonesia
Phone: +62-21-546-0901 ext.1348
arnold.aribowo@uph.edu

ABSTRACT
This paper proposed a modification of paper-printing printers for garment printings usually called Direct-to-Garment (DTG) printing. This method is suitable for low quantity production and even unique product, namely an arbitrary image printed on a single garment. The modified device, based on Epson printers, has been tested both in software and hardware.

Keywords
printer modification, direct to garment printing.

1. INTRODUCTION
Ordinary printers, usually, are only capable of printing on papers or paper-based mediums. Normally, they have to print black and white or colors on white. Other background colors may apply, but it would certainly affect the final outcome of the printing process.

Some of these printers are equipped with combined-colors cartridges, namely one cartridge with three (RGB) or four (CYMK) main colors, and some others are equipped with separate cartridges for each color. The main colors are then combined, when printing, with a certain composition to produce certain desired colors, as faithfully as possible to the printed image. Some more advanced printers are even equipped with two cartridges per main color, e.g. cyan and light cyan, to be able to achieve certain color depths in its mission to imitate its pattern image.

On the other hand, garment printing, i.e. Direct-to-Garment (DTG) printing, is a fast-growing business in today's market. Giant entrepreneurs are able to compete with big funding to install sophisticated printing machines for their mass production activities. However, the entrepreneurs of lower budget need to find other alternatives in order to be able to join the cruel and tough competition. In this situation, we have come up with a modest and yet robust solution for garment printing.

The ordinary printers existing in the market are modifiable to do garment printing. It may not be as fast and reliable as their high-end counterparts owned by big garment factories, but it indeed has certain comparable advantages. Hence, we tried to modify this paper-based printing device into garment printing device. This paper discusses this solution including the design, implementation, and some future thoughts regarding this experimentation in the next sections.

2. RELATED WORKS
Garment printing is already a widespread, although it is new, in textile printing domain. The technology of digital printing described in [2] gives some insights on how digital printing technology emerges from the existing printing devices. In [3], many alternatives and insights are explored on how digital printing techniques can foster the creativity of the designers and ease their efforts in putting their designs on clothing.

The existing products comprise a diverse range of alternatives. The screen printing explained in [5] is a well-established solution and quite complex. Another DTG solution in [6] is also based on Raster Image Processing (RIP) which is also still in the digital printing mainstream provides more advanced printer driver manipulation to create a robust digital printer. Some existing DTG printers are even manufactured and mushrooming as explained in [4]. However, we are trying to start another alternative research from the beginning, in collaborative work with the industry, i.e. lower-budget entrepreneur, to create an alternative digital garment printing solution which is robust, cheaper, and quite flexible for further development, while taking advantage from the existing open-source resources.

3. DESIGN AND IMPLEMENTATION
3.1 Design Considerations
The printer should be modified in a way that would enable garment printing. The inner parts of the printer casing, namely the printing machine, e.g. electronics, motor, head, etc., are to be placed on a new platform specialized for garment printing. In this case, the paper tray is replaced with something else. The cartridges are also replaced with those specialized for textile inks. The mechanical parts, e.g. gears, belt, etc., have to be restructured on the new platform in order that the printer would function normally. In general, the transformation taking place here is to change, namely, a paper printer into a garment printer.

The printers used for this garment printing purpose are from the *Epson* printer family. They are used because of flexibility and modifiability, but yet reliable and robust. The printer cartridges can still function very well when they have to be replaced by those specialized for textile inks. The printer mechanical components are also reliable when ported into the new platform.

3.2 Hardware Design
The complete design of the printer placed on the new platform specialized for this garment printing is shown in Figure 1 [1].

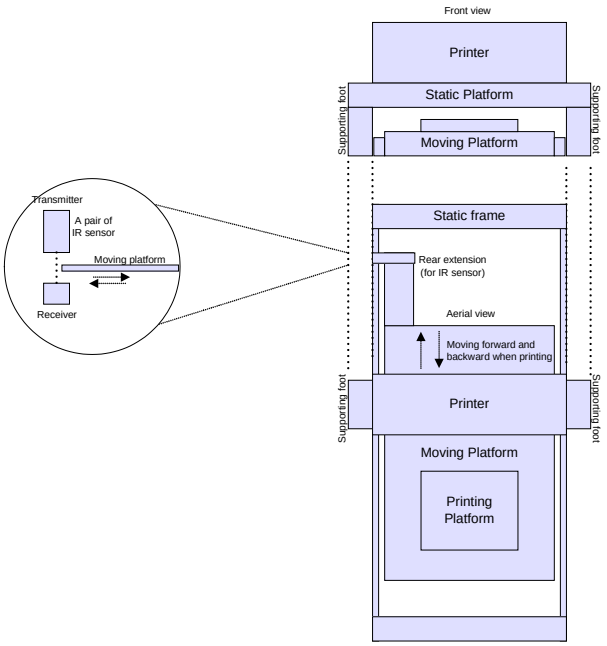


Figure 1. Modified printer hardware (aerial view, front view, IR sensor) [1]

Aside from the new platform for the printing machine, another major hardware modification is the additional controller for the printer's DC motor [1]. By using this additional handy controller, a user can actually operate the motor as he might want to move the printing platform backward an forward. There are three modes of movement: forward, backward, and back to ready position (based on the IR sensor placed at the back of the static frame of the platform). The controller consists of some components: *Latih-Xtra* module with AVR processor on top of it, a 4 x 4 keypad matrix, an IR sensor circuitry (shown in Figure 2), and a printer DC motor driver (shown in Figure 3).

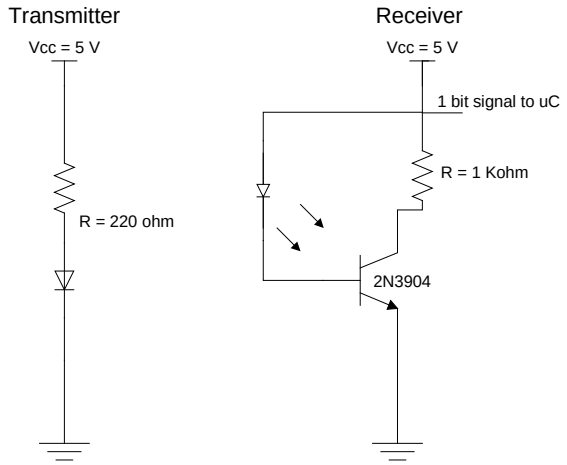


Figure 2. Schematic diagram of IR sensor circuitry

The IR sensor circuitry consists of two parts, namely *transmitter* and *receiver*. The transmitter part consists of a resistor and a LED that emits infrared (IR) light spectrum. The receiver part consists of a resistor, an LDR, and a transistor. The LDR is to detect, whether or not the LED emits IR light. It then sends a 1-bit signal to indicate the reception of the IR light.

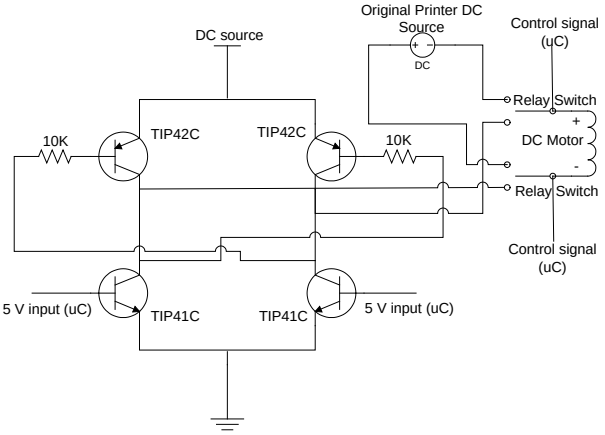


Figure 3. Schematic diagram of DC motor driver

The DC power source for the printer DC motor is regulated using an H-bridge circuitry built from a pair of PNP transistors and another pair of NPN transistors. It is the main part of this motor driver.

Another important part is two relays added to the circuit for choosing between the original printer source and the external DC power source (controlled by H-bridge). Putting two electricity sources into one point is considered dangerous to the circuit. The complete schematic of the complete DC motor controller system is shown in Figure 4 and the flowchart of the code running on the AVR processor controlling the three movement of the printing platform is shown in Figure 5.

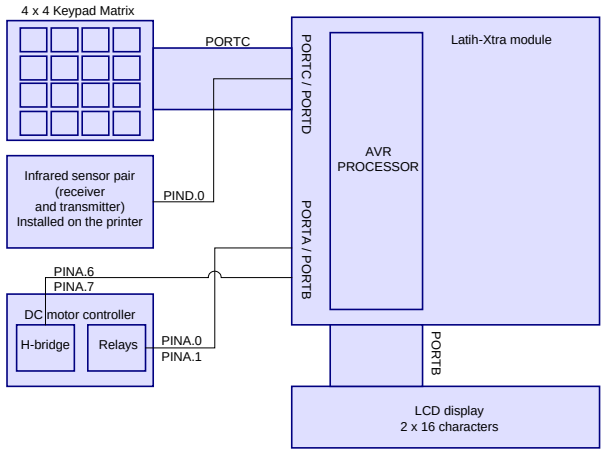


Figure 4. Schematic diagram of complete DC motor controller system

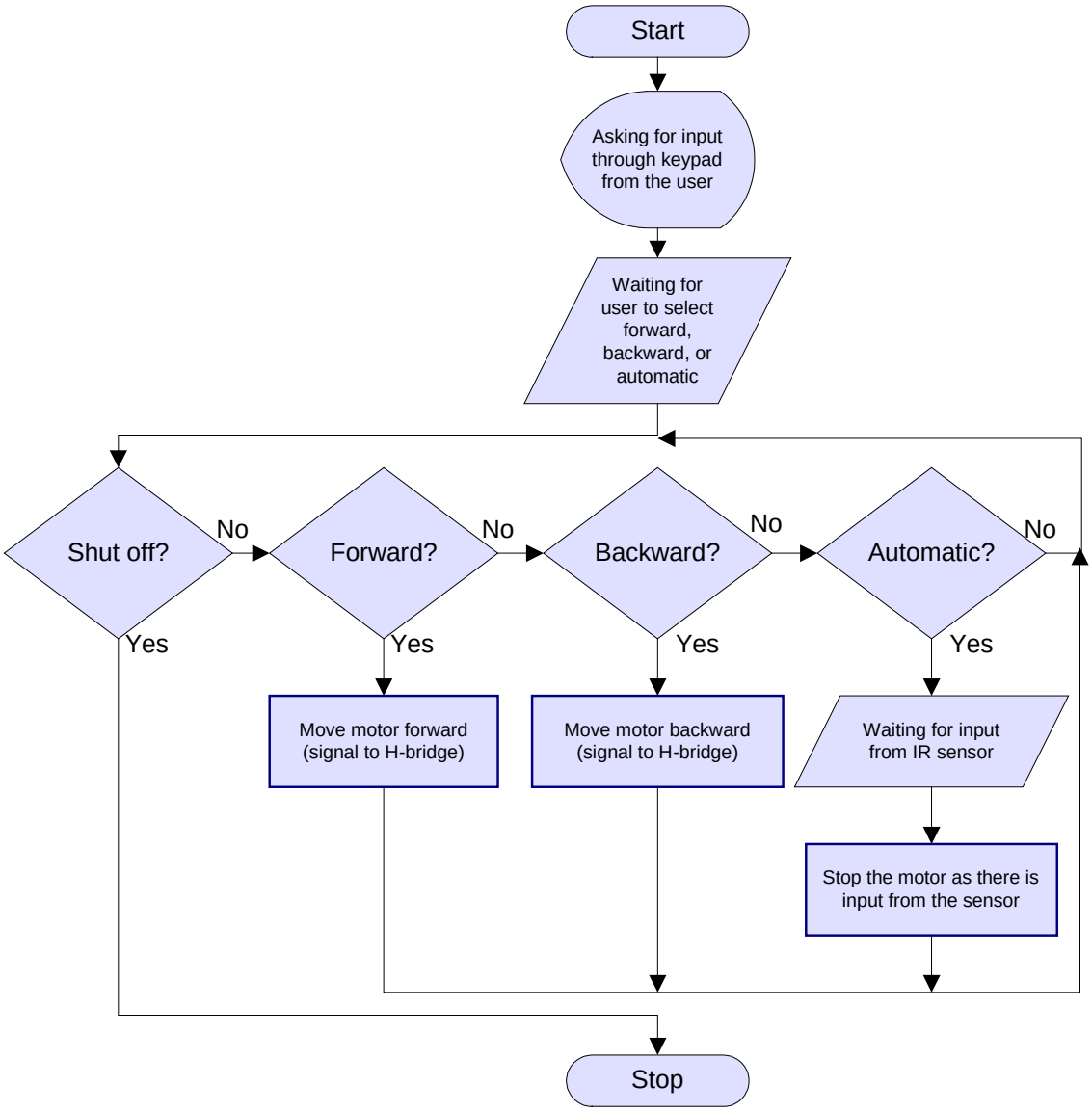


Figure 5. AVR software flowchart

3.3 Software Design

The software is also created to run on PC/laptop. The operating system used during the experiment is Ubuntu 10.10 (Maverick Meerkat) from Linux open-source operating system family. The specific software used for printing is Gutenprint. Gutenprint is actually an open-source universal printer driver that accommodates a lot of printer types for UNIX and Linux, including MacOS X, operating systems.

The condition that has to be achieved in garment printing is to be able to print images on garments. White-colored garments are not a problem for printers in general. Ordinary printer drivers are able to spray right inks for various colors through specific printer heads. In this case, images would be printed well.

A big problem, however, occurs if we are to use garments with other colors. Inks with different colors would not successfully reconstruct an image on garment, because the color of the inks would blend into the color of the garment. In this case, we need to create a white foundation on the garment before printing the colorful image on top of it. This way, the image would be printed well according to the real image colors. Therefore, the ordinary printers are to be modified to achieve this purpose.

There are two things to achieve: printing white foundation and printing colorful image. The printer driver software should be able to print both white foundation and image. It should give an option

for users to choose between the two. The users would then be able to choose to print the white foundation, as much as it is needed, or to print the image. Normally, the white foundation should come first before the image. The source code from Gutenprint is modified and recompiled to achieve this purpose. The flowchart for this software is shown in Figure 6.

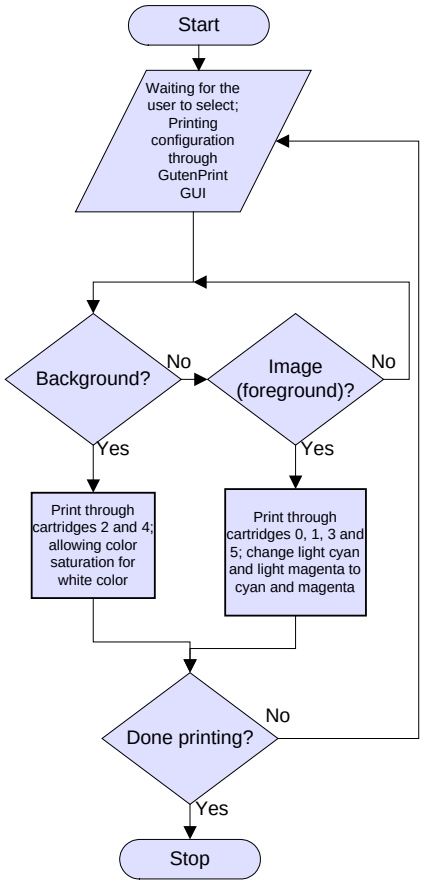


Figure 6. Software driver flowchart

3.4 Results

The printing results are acceptable and even satisfactory. Figure 7 shows the result of printing white foundation on a non-white garment. Figure 8 shows the result of printing a white foundation on non-white garments and a colorful image on the white foundation.

Figure 7. Result of printing white foundation on non-white garment

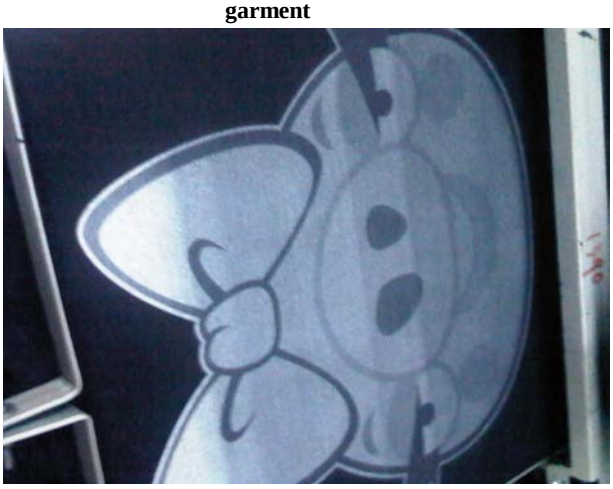


Figure 8. Result of printing white foundation (left) and colorful image on top of white foundation (right)

4. CONCLUSIONS AND FUTURE WORKS

In this paper, we have shown how an alternative garment printing device is designed and implemented. The device is to achieve the purpose of printing images on garments. The printing device is originally an ordinary printer modified to accomplish this need.

Basically, ordinary printers can print inks to form an image on a medium. However, it is limited to white-colored media. Images printed for non-white media would suffer from color distortion due to the background color. Our device has been tested and proven to achieve this need by facilitating white ink printing as a foundation before printing the image.

This device, however, has some drawbacks. It cannot print the white foundation and the image in order automatically. Users have to control the order manually. The white ink is also considered too thin, so a greater amount of ink should be printed from the printer heads, in order not to repeat the white ink printing for several times. This research can also be extended to other types of printers.

5. ACKNOWLEDGMENTS

We would like to thank University of Pelita Harapan and Indo.co.id, a garment printing company, that have supported this valuable research as a modest alternative to the existing solutions.

6. REFERENCES

[1] Trimananda, R. 2011 Research on Garment Printing. Technical Report. University of Pelita Harapan.

[2] Cahill, V., "Introduction to Digital Printing Technology", Journal of Specialty Graphic Imaging Association,

[3] Campbell, J.R., Parsons, J., "Taking Advantage of The Design Potential of Digital Printing Technology for Apparel", Journal of Textile and Apparel, Technology and Management, 2005

[4] Flower, G. 2011 What's New in Direct-to-Garment Printers? OmniPrint International. http://www.omniprintonline.com/articles/Direct_to_Garment_Printing_Whats_new.pdf

[5] Sefar. 2011 T-Shirt/Garment Printing. Product Detail. Sefar. http://www.sefar.com/htm/609/en/Product-Detail.pdf?Folder=1465804

[6] MultiRIP. 2011 http://www.multirip.com/index.html.

[7] Krawitz, R. 2006 Gutenprint 5.0 User's Manual. Gutenprint.

[8] The Gutenprint Project. 2003 The Developer's Guide to Gutenprint. Gutenprint.

Authors Index