Jurnal Teknik Industri, Vol. 17, No. 1, Juni 2015, 47-52
ISSN 1411-2485 print / ISSN 2087-7439 online
DOI: 10.9744/jti.17.1.47-52

# Sequence Matching Analysis for Curriculum Development

**Liem Yenny Bendatu[1]*, Bernardo Nugroho Yahya[2]**

**Abstract**: Many organizations apply information technologies to support their business processes. Using the information technologies, the actual events are recorded and utilized to conform with predefined model. Conformance checking is an approach to measure the fitness and appropriateness between process model and actual events. However, when there are multiple events with the same timestamp, the traditional approach unfit to result such measures. This study attempts to develop a sequence matching analysis. Considering conformance checking as the basis of this approach, this proposed approach utilizes the current control flow technique in process mining domain. A case study in the field of educational process has been conducted. This study also proposes a curriculum analysis framework to test the proposed approach. By considering the learning sequence of students, it results some measurements for curriculum development. Finally, the result of the proposed approach has been verified by relevant instructors for further development.

**Keywords**: Sequence matching analysis, process mining, curriculum analysis, learning sequence.

## Introduction

Using information technologies in many organizations, including educational field, are growing in order to support their respective processes. The utilization of these systems streamlines the processes and records the actual events in the repository which often called as events log (Rozinat, and Aalst [1]). Generally, the event log consists of actual data with related timestamp and can be further analyzed to construct an as-is process, i.e., the actual process. Using relevant approaches, the constructed result fits for comparison with the predefined model. However, it raises a question 'is the actual process representing the predefined model'.

Some works have been done to answer the aforementioned question. Among the works, conformance checking is considered the suitable approach to measure the gap between actual process and predefined model (Rozinat, and Aalst [1], Aalst [2]). However, the approach has limitations to handle multiple events with the same timestamp. For example, in an education institution, different courses can be taken in the same semester. Therefore, it is necessary to improve the traditional approach for curriculum problem.

[1] Faculty of Industrial Technology, Industrial Engineering Department, Petra Christian University, Jl. Siwalankerto 121-131, Surabaya 60236. Email: yenny@petra.ac.id
[2] Industrial and Management Engineering Department, Hankuk University of Foreign Studies, Oedaero 81, Mohyeonmyon, Cheoingu, Yongin, South Korea 449791.
Email: bernardo@hufs.ac.kr

* Corresponding author

The conformance checking for multiple events with the same timestamp, as later called by sequence matching analysis, needs to consider several relevant attributes such as activity and timestamp. One of the challenges to conduct sequence matching analysis is to organize the same timestamp into a group. Regardless the order of events, activity with the same timestamp will be considered in the same group of time. In this sense, it requires specific approach to analyze these particular events.

By partially reusing the control flow approaches in process mining, this work aims to develop a new technique of sequence matching analysis to measure educational process. There was an attempt on curriculum mining. However, it has less coverage on conforming the learning sequence to curriculum model. This study utilizes student's performance data of an engineering program as the input for the evaluation. The final result of this work will benefit not only the student's performance but also the program performance. In addition, it will further help instructors to refine and improve the curriculum. In a broad perspective, this technique can be used by any educational programs to conform their curriculum in term of educational process.

### Related Work

A curriculum is a term refers to academic contents in a specific program brought by educational institutions. Typically, a curriculum refers to knowledge and skills students are expected to learn, which includes learning standards and learning process they are expected to meet. Shao-Wen Su [3] mentioned that curriculum have several definitions;

they are mentioned as a set of objectives, courses of study or content, plans, documents or even experiences. Pratt ([4] p.5) and Barrow and Milburn ([5], p. 84) presented a curriculum, in Latin "currere", as a running track where the track shows a distance to be achieved by a runner from the start to the finish line. Goodson described curriculum as a multifaceted concept, constructed, negotiated, and renegotiated at a variety of levels and in a variety of arenas" ([6], p. 111). This showed that a curriculum is naturally complex and interactive. In addition, Longstreet and Shane [7] showed that a curriculum is a tool for decision maker. Therefore, a curriculum can be regarded as a collection of courses used as the reference both students and instructors in educational process to achieve competencies of students' profiles.

Related work on conforming the actual process and the respective model has been done by Rozinat and Aalst [1], Aalst [2]. The proposed technique, called Conformance Checker, demonstrated the fitness and the appropriateness between the process model and event log (Rozinat and Aalst [1]). The work discussed the potential used of conformance checking to locate the misalignment between the process model and actual behavior which is from the information system (Aalst, [2]).

Related work on curriculum mining in education institution has been done by Pechenizkiy *et al.* [8]. He was evaluating based on a case study from students at Eindhoven University of Technology. According Pechenizkiy *et al.*, an academic curriculum is a legal document defining a specific learning program that puts certain types of constraints on how students are expected to take the courses (Pechenizkiy *et al.* [8]). Students can choose what subjects they want to take for each semester and lead them to have different learning sequence. While the rules usually stated informally in the current practice, this will lead to multiple interpretations for students and educators. There were 3 main results from the previous work. First, it discovered the actual model behavior. Second, checking the model whether the expected behavior is matched with the actual model. Lastly, the curriculum model extension is to make the tacit knowledge explicit or to help the educators on decision making process.

To effectively analyze the curriculum, there are several analysis works using data mining. According to Witten and Frank [9], data mining is the process oriented to extract useful and comprehensible knowledge, previously unknown, from huge and heterogeneous data repository. Data mining is data analysis methodology used to identify hidden patterns in a large data set of a curriculum (Chandra, and Nandhini [10]). Process mining captures the data in sequence which means that every activity has timestamp. Process mining has emerged as a way to analyze system and their actual use based on the event logs they produce (Aalst, [11]). Process mining has been used to model the existing academic curriculum with formal language of Colored Petri nets (CPNs). The data, called event log, will be processed through process mining to provide useful knowledge such as process models, organizational models, etc (Aalst [12,13]).

This study attempts to evaluate other format of education curriculum and uses a case study of an engineering department at an education institution in Indonesia. The proposed approach in this study is somewhat different with the previous approach (Pechenizkiy *et al.* [8]). In Indonesia, the students do not have the freedom to choose each subject like the practice in the previous work. Some courses have been set in each semester and some other courses have been provided to be chosen by students. Although the courses have been set into each semester, the students can take those subjects earlier or after the assigned semester following with prerequisite of the courses. This paper contributes to evaluate the a-prori curriculum design with actual curriculum learning sequence through sequence matching analysis.

## Methods

### Research Methodology

The development of sequence matching analysis in the domain of curriculum requires several steps for verifying the evaluation. This study proposes a framework of curriculum analysis for testing the sequence matching analysis. In the later part, the term matching analysis will be used for sequence matching analysis.

The framework of curriculum analysis consists of three steps, data preparation, curriculum analysis and verification. Data preparation aims to extract relevant data from database and preprocess for the curriculum analysis steps. The curriculum analysis step aims to evaluate the students' learning process correspond to curriculum model. Finally, the verification step aims to measure the fitness of the curriculum analysis method. The proposed curriculum analysis framework is shown at Figure 1.

In the information systems, there are a lot of data stored in the database. For this study, relevant data with regard to student's performance are extracted. Some relevant data used in this study are student personal data, courses taken each semester, number of credit as per course, and grade of the course.

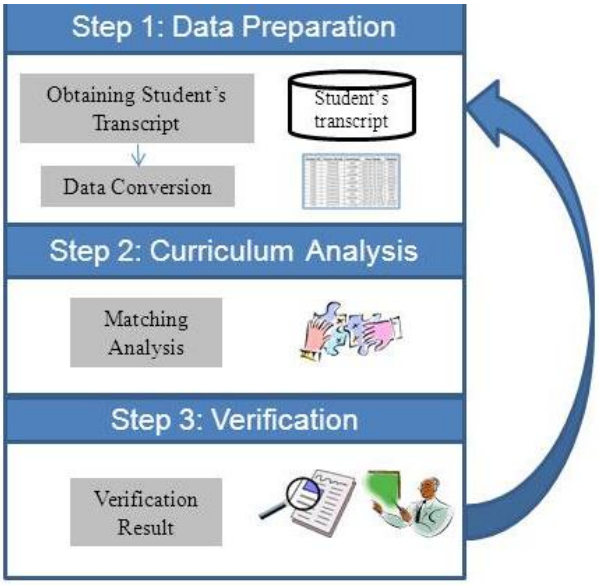


**Figure 1.** Curriculum analysis framework

The data extracted from the information system may consist of incomplete data, for example student who is still doing the study or transferred student.

Therefore, the data used in this study should be filtered to only data with regard to students who have completed their study and experience the learning process from the 1st semester. Afterward, it is necessary to convert the extracted data to a formal form for the curriculum analysis. In this study, the data is converted to MXML file.

From the curriculum analysis, it is necessary to formulize the data as follow. In this study, log and events are defined as follows.

A log is a tuple of ($S$, $E$) where $S$ is a set of student ID and $E$ is the set of events. $E$ is a set of < $\hat{C}$, $G$, $\check{T}$> where $\hat{C}$ is the set of students' data courses, $G$ is the grade of the related course and $T$ is the period that the course is taken. The notion of $e_1.\hat{C}$, $e_1.G$, and $e_1.\check{T}$ are used to denote a course name, the grade and the semester period, respectively. For example, $e_1$ = {***A***, B, 1-09/10} means that a course ***A*** with grade B was taken at the 1st semester of 2009/2010. Hence, $e_1.\hat{C}$ ="***A***", $e_1.G$="B", and $e_1.\check{T}$="1-09/10".

The fragment of the data is shown at Table 1. It shows three students' data from 1st semester to 4th semester with some courses in the final semester. The courses shown in the table are as follows; Introduction to Industrial Engineering (***A***), Calculus I (***B***), Basic Computer and Programming (***C***), Calculus II (***D***), Matrix and Vector Space (***E***), Advance Calculus (***F***), Database (***G***), Export Import Management (***H***), Technopreneurship I (***I***). The fragment data from the a-priori curriculum model is shown in Table 2.

**Table 1.** Fragment data from student's transcript

| Student | Course | Grade | Semester |
|---|---|---|---|
| 1 | ***A*** | E(Failed) | 1-09/10 |
| 2 | ***A*** | B | 1-09/10 |
| 3 | ***A*** | C+ | 1-09/10 |
| 1 | ***B*** | D | 1-09/10 |
| 2 | ***B*** | D | 1-09/10 |
| 3 | ***B*** | C | 1-09/10 |
| 1 | ***B*** | C+ | 2-09/10 |
| 2 | ***B*** | C | 2-09/10 |
| 1 | ***C*** | E(Failed) | 2-09/10 |
| 2 | ***C*** | B | 2-09/10 |
| 3 | ***C*** | C+ | 2-09/10 |
| 3 | ***D*** | E(Failed) | 2-09/10 |
| 1 | ***D*** | B | 1-10/11 |
| 2 | ***D*** | C | 1-10/11 |
| 3 | ***D*** | C+ | 1-10/11 |
| 3 | ***E*** | C+ | 1-10/11 |
| 1 | ***F*** | C | 2-10/11 |
| 2 | ***F*** | D | 2-10/11 |
| 3 | ***F*** | D | 2-10/11 |
| 2 | ***G*** | B | 2-10/11 |
| 3 | ***G*** | B | 2-10/11 |
| 1 | ***H*** | B+ | 1-12/13 |
| 2 | ***I*** | B | 2-11/12 |

**Table 2.** Fragment data from the curriculum model

| Course | Predefined Semester |
|---|---|
| ***A*** | 1 |
| ***B*** | 1 |
| ***C*** | 1 |
| ***D*** | 2 |
| ***E*** | 3 |
| ***F*** | 4 |
| ***G*** | 4 |
| ***H*** | 6 |
| ***I*** | 6 |

To evaluate the students' performance, matching analysis is used. Matching analysis is used in many fields based on statistics measures. For example, there are two runs with treated and non-treated units that need to be analyzed. Paired difference test can be used to measure the average treatment effect. Matching analysis can also be used to match the learning sequence from the timestamp of the predefined model with the actual model [1]. This study presents the sequence matching analysis to test the period difference between the curriculum model and the student's historical data. To differentiate the course of curriculum model and students' data, let $C$ be the set of courses in curriculum model. It should be noted that the curriculum model has also attributes of prerequisite grade. This study will use only the predefined time and put the prerequisite grade for future work. To match the time period, there should be a function to convert the student's data period into relative time. A function $ft$ is used to convert time of $e.\check{T}$ into a relative time. For example, if $e.\check{T}$="1-09/10" is the

first semester of the student, then *e.T*="2-09/10" is the second semester, *e.T*="1-10/11" is the third semester, and so on. Let $c$ be the course in the curriculum model and $\hat{c}$ be the course taken by a student. Then $ft(c)$ is the relative time of the course $c$ in the curriculum model and $ft(\hat{c})$ be the relative time of the course $\hat{c}$ taken by the student. The sequence matching analysis between the relative time course in the curriculum model and the relative time course taken by the student is formulized as follow.

$$m_s(c,\hat{c}) = \begin{cases} 1 & \text{if } c = \hat{c} \text{ and } f_t(c) = f_t(\hat{c}) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Overall sequence matching analysis $M$ of a course aims to measure all students' performance with respect to a specific course. The result will be a range between 0 and 1. A result of 1 means that the course is taken exactly at the same predefined. A value nearby one gives an intuitive result that the student's learning process matches with the curriculum model. Hence, the frequency measure of matching (*FM*) and relative measure of matching (*RM*) are formulized as follow.

$$FM(c,\hat{c}) = \sum_{s=1}^{S} m_s(c,\hat{c}) \quad (2)$$

$$RM(c,\hat{c}) = \frac{FM(c,\hat{c})}{S} \quad (3)$$

For example, $m_1$ (***A***, ***A***) =1 since the student 1 took course ***A*** in the same semester period as the predefined semester. However, $m_2$ (***D***, ***D***)=0 since the student 2 took the course ***D*** not in the same period of the predefined. The frequency measure of course ***A***, (*FM*(***A***, ***A***)) is 3 and the relative measures of course ***A*** (*RM*(***A***, ***A***)) and course ***D*** (*RM*(***D***, ***D***)) is 1 and 0.33, respectively.

In fact, a student can take the same course several times due to failing to satisfy the prerequisite. Herein, it is called by retaking course. For example, student 2 took course ***D*** two times. Student 2 failed to qualify the course ***D*** requirement and retook the course ***D*** in the next semester. Since the intention is to measure the matching between the student's performance and curriculum model, this study disregards retaking courses and measures only courses taken at the first time by students.

A course taken by the student can be in the same period with, before or after the period of the curriculum model. The course is regarded as before the period of the curriculum model when the specific course is taken before the relative time of the specific course. For example, the curriculum model set course ***I*** to be taken on semester 6 and the student took course ***I*** in the semester 5, which is earlier than the predefined period. Using the same equations, the courses taken either before or after the predefined period can be also measured. The formulization of before_period ($bm$) and after period ($am$) are as follow.

$$bm_s(c,\hat{c}) = \begin{cases} 1 & \text{if } c = \hat{c} \text{ and } f_t(c) = f_t(\hat{c}) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$am_s(c,\hat{c}) = \begin{cases} 1 & \text{if } c = \hat{c} \text{ and } f_t(c) = f_t(\hat{c}) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Therefore, the frequency measures for before ($FBM$) and after ($FAM$) are formulized as follows.

$$FBM(c,\hat{c}) = \sum_{s=1}^{S} bm_s(c,\hat{c}) \quad (6)$$

$$FAM(c,\hat{c}) = \sum_{s=1}^{S} am_s(c,\hat{c}) \quad (7)$$

In addition, the relative measures for before ($RBM$) and after ($RAM$) period are formulized as follows.

$$RBM(c,\hat{c}) = \frac{FBM(c,\hat{c})}{S} \quad (8)$$

$$RAM(c,\hat{c}) = \frac{FAM(c,\hat{c})}{S} \quad (9)$$

The result of curriculum analysis using sequence matching analysis will be verified by the instructors of the related program. The instructors could check the measures and find the problems on the current curriculum model for further development.

## Results and Discussions

This section aims to show the implementation results and discuss some relevant issues of the current work. First, the data used in this study is described. Second, the tool used for curriculum analysis is explained. Finally, instructor's verify-cation is presented in the discussion section.

This study applies a case study from an engineering department in one university in Indonesia. The data used in this study consists of students which used curriculum 2008 as the model. The data which consists of noises has been filtered. For example, data which consists transferred students has been ignored. For the experiment of the curriculum analysis, there were 39 students' performance data as the result of data filtering phase. The data is converted to MXML for further analysis.

For the analysis, this work utilizes an open source process mining tool, called ProM [14]. The tool is regarded as a supporting tool for developing the matching analysis. The tool has some prominent features for this study. First, the student's learning sequence can be regarded as a trace in the process. Second, the use of timestamp in process mining is the same as the intuitive idea of this work. The related plugin with regard to timestamp is reused for this study as the basis on developing matching analysis approach. The result of the study is a visualization in the form of table. In addition, for further verification, some instructors of the related department have been chosen to analyze the result.

## Implementation Results

The proposed approach has been developed in a plugin under ProM tool [14]. The result in the table format can be seen in Figure 2. It shows the snapshot of the matching analysis result. It has the information on each columns; semester, the list of courses, *FBM* (Frequency Before Matching), *RBM* (Relative Before Matching), *FM* (Frequency Matching), RM (Relative Matching), *FAM* (Frequency After Matching) and *RAM* (Relative After Matching).

It can be observed that the core engineering subjects in 1st semester are having the *FM* and *RM* 39 students and 1, respectively. It means there are 39 students taking those subjects exactly as in the a-priori model, and the *RM* is from 39/39. This is due to the compulsory subjects that all students have to go through in their 1st semester.

Unlike the data of the semester 1, the result of data from semester 2 up to semester 8 shows that not all subjects have been taken exactly as the predefined curriculum model. The matching result (*RM*) varies with the range value between 0 and 1. As a consequence, there could be some values on *RBM* and *RAM*. Finally, we could check that $RBM + RM + RAM = 1$.

There are several courses have been taken earlier than the predefined curriculum. For example, courses with relative result more than 0.66 is *Introduction to Economics and Cost Analysis*. While subjects have been taken after than a-priori model are *Industrial Statistics*, *Statistics for Experimental Designs* and *Statistical Quality Control* with relative result more than 0.5.

**Matching Analysis View**

| Semester | Course | FBM | RBM | FM | RM | FAM | RAM |
|---|---|---|---|---|---|---|---|
| 1 | ETHICS | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | CALCULUS I | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | RELIGION PHILOSOPHY | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | CIVIC AND IDEOLOGY | 0 | 0.0 | 8 | 0.2051 | 31 | 0.7949 |
| 1 | BASIC PHYSICS I | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | ENGLISH FOR ACADEMIC PURPOSES | 0 | 0.0 | 31 | 0.7949 | 8 | 0.2051 |
| 1 | ENVIRONMENTAL KNOWLEDGE | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | INTRODUCTION TO INDUSTRIAL ENGINEERING | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 1 | SCIENTIFIC WRITING AND COMMUNICATION | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 2 | BASIC PHYSICS II | 0 | 0.0 | 39 | 1.0 | 0 | 0.0 |
| 2 | CALCULUS II | 0 | 0.0 | 37 | 0.9487 | 2 | 0.0513 |
| 4 | DATABASE | 0 | 0.0 | 34 | 0.8718 | 5 | 0.1282 |
| 4 | INTRODUCTION TO ECONOMICS | 26 | 0.6667 | 10 | 0.2564 | 3 | 0.0769 |
| 4 | INTRODUCTION TO BUSINESS AND MANAGEMENT | 0 | 0.0 | 35 | 0.8974 | 4 | 0.1026 |
| 4 | ENGINEERING PSYCHOLOGY | 0 | 0.0 | 20 | 0.5128 | 19 | 0.4872 |
| 4 | ADVANCED CALCULUS | 19 | 0.4872 | 17 | 0.4359 | 3 | 0.0769 |
| 5 | COST ANALYSIS | 26 | 0.6667 | 10 | 0.2564 | 3 | 0.0769 |

**Figure 2.** Snapshot of the matching analysis result in the semester 4.

Using the same measurement, we could analyze the learning sequence for other batches and related curriculum model which might help the educators to see the result from the past curriculum model. Hence, this matching analysis result will give contribution to the program on evaluating and developing the curriculum.

## Discussions and Verifications

Based on the analysis result, there are some interesting findings. Those findings were analyzed by designated instructors. The instructors verify the result and inform the result to process analyst in corresponding to each measurement.

Some interesting findings are related to courses that have been taken not in the predefined semester period. Using the *RBM* and *RAM* measurement, it could be checked whether the students took the courses before or after the predefined semester. Since the value of *RBM* is greater than *RAM*, it is believed that most students took the courses before the predefined semester. The highlighted result is the matching rate that equals to a certain threshold which is minimum 0.5. First, there are 3 subjects out of 12 subjects that has *RBM* value more than 0.5. Two subjects are the supporting courses and 1 subject of core course. One of the supporting subjects shown in figure 2 is *Introduction to Economy* which has been taken earlier than the predefined semester. This is due to there is no pre-requisite on this subject, hence the students would like to take it earlier as they still have the total credits remain.

Second, there are 36 subjects out of 49 subjects that have *RM* value more than 50%. Eighteen subjects are the supporting courses, 7 subjects are the basic subjects while 11 subjects are the core subjects. One of the supporting subjects shown in figure 2 is *Environmental Knowledge* has been taken precisely with the predefined curriculum model. Most subjects that have been taken precisely were taken in the early semester, such as completely in semester 1 and some in semester 2-4.

Lastly, there are 6 subjects out of 35 subjects that have *RAM* value more than 0.5. Two subjects are the supporting courses, 3 subjects are the core courses, while 1 subject of basic course. The example core course is Statistical Quality Control which has been taken after predefined curriculum model. This is due to the pre-requisite of the subject which requires the students to pass a sequential subject which is *Probability Theory* and *Industrial Statistics*.

Based on the above result, the curriculum applied has shown that many subjects have been taken after the predefined curriculum model, especially after the 4th semester. This result will be given as the feedback to the program.

Another finding is the existence of certain course that is not in the curriculum model such as *Technoprenurship II*. This is related to the curriculum development issue. It is because of some students took courses when the curriculum has been changed. The latest curriculum is 2012 whereby the students from batch 2009 experienced the changes in their learning period.

## Conclusion

This study aims to develop sequence matching analysis for multiple events with the same timestamp. The approach was tested to analyze the curriculum using student's performance data. It utilized process mining tool, called ProM, for analysis. Since there is no relevant plugin for the curriculum analysis, a new plugin had been developed. This plugin could classify multiple events with the same timestamp to indicate the sequence of students' learning process. In addition, it measured the difference between predefined periods with actual learning period of students.

There were some findings on the current work. First, it discovered that some courses have been taken not exactly the same as the predefined period. Second, some of the courses were overlapped with the new curriculum proposed in the department.

There are some issues opened for future work. Enlarging the data from batches that experience the same curriculum model will give more comprehensive result. Moreover, the existing analysis can also be combined with other measures to find the bottleneck of students' learning process. Other issues are related to attributes such as gender and students' historical background. The relationship between the learning process and those attributes could be an input factor for curriculum development.

]

## References


1. Rozinat, A., and Aalst, W.M.P., Conformance Checking of Processes Based on Monitoring Real Behaviour, *Information System* 33, 2008, pp 64-95.
2. Aalst, W.M.P., Business Alignment: Using Process Mining as a Tool for Delta Analysis and Conformance Testing, *Springer* 2005.
3. Shao-Wen, Su., The Various Concepts of Curriculum and the Factors Involved in Curriculamaking, Journal of Language Teaching and Research, 3(1), 2012, pp. 153-158.
4. Pratt, D., *Curriculum Planning: A handbook for professionals*. Fort Worth: Harcourt Brace College Publishers, 1994.
5. Barrow, R., and Milburn, G., *A Critical Dictionary of Educational Concepts*. New York: Harvester Wheatsheaf, 1990.
6. Goodson, I. F., *Studying Curriculum*. New York: Teachers College Press, 1994.
7. Longstreet, W.S. and Shane, H.G., *Curriculum for a New Millennium*. Boston: Allyn and Bacon, 1993.
8. Pechenizkiy, M., Trcka, N., De. Bra, P., *CurriM: Curriculum Mining*, EDM 2012, pp. 216-217.
9. Witten, I.H., and Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Fransisco, 2000.
10. Chandra, E., and Nandhini, K., Knowledge Mining from Student Data. European Journal of Scientific Research, 47(1), 2010, pp. 156-163.
11. Aalst, W.M.P., Process Mining: Discovery, Conformance and Enhancement of Business Processes, *Berlin: Springer-Verlag*, 2011.
12. Aalst, W.M.P., Weijters, A., and Maruster, L., Workflow Mining: Discovering Process Models from Eveng Logs, *IEEE Transactions on Knowledge and Data Engineering*, 16(9), 2014, pp. 1128-1142
13. Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Aalves de Mediros, A.K., Song, M.S., and Verbeek, H.M.W., Business Process Mining: An Industrial Application, *Information System* 32, 2007, pp. 713-732.
14. Pro M, *Process Mining Tools*, Available: http:// www.processmining.org, retrieved on April 2014