

Web & Android Application for Harvester of Indonesian Scientific Paper Citation

Resmana Lim

Electrical Engineering
Petra Christian University
Surabaya, East Java, Indonesia
resmana@petra.ac.id

Raymond Sutjiadi

Research Center
Petra Christian University
Surabaya, East Java, Indonesia
raymondsutjiadi@petra.ac.id

Adi Wibowo

Informatics Engineering
Petra Christian University
Surabaya, East Java, Indonesia
adiw@petra.ac.id

Abstract—This research proposes a harvester system that utilize Indonesian language based parser to capture citations' metadata from papers. Open Harvester System from Public Knowledge Project (PKP-OHS) is used as a base of harvester system. Citation extraction and citation citegraph methods are added to extend the processing capability of PKP – OHS to enable processing citations. Several information output are modified to enable provision of citation information to users. This system is also equipped with an Android application to give the paper access through mobile devices.

Keywords - harvester system, citation extraction, android application

I. INTRODUCTION

Service and application software utilizing mobile technologies such as the smartphone and pc-tablet (ipad /android) has penetrated into various field ranging from entertainment, business, health, lifestyle and do not miss the education. Convergence of computing technologies, multimedia content and telecommunications has changed the lifestyle of many included in the way of scientific community accessing scientific papers through their mobile device. Some mobile applications like Scholay and Mobile Mendeley deliver scientific paper access for mobile users. This application runs on Android, gives access to paper's abstracts of existing papers on Mendeley.com. However the user can not obtain the paper's citations. Users only get related papers from mendeley.com [1].

This research proposes a scientific paper's citation Web access and additional access via mobile devices (Android Application). In Indonesian context, there is a lack of support in Indonesian scientific repository to provide complete access to Indonesian papers and journals. Usually each repository will maintain each own database, and also its own access methods. There is also still no citation network between papers inside a repository and also between papers among repositories.

This research proposes a harvester system that is part of a larger research project to develop Indonesian Scientific Citation Database (ISCD) System [2]. ISCD system try to

provide a database system consists of Indonesian papers and journals from Indonesian researchers, and also to build citation network among papers, and citation analysis that analyse researcher's and journal's impact factor, citation statistics, topic of interest, and other metrics. To achieve this goal the project can not run in solitary but must utilize scientific article databases held by many institutions in Indonesia as content providers. The ISCD will harvest articles' metadata from several journal database in Indonesia, parse each article's PDF files, store and then link citation to the original article. The server also provide the web services for Android application to access the paper's information & citation from mobile devices (smartphones/pc-tablets), based of Android operating system. The overview of the project is shown at Figure 1.

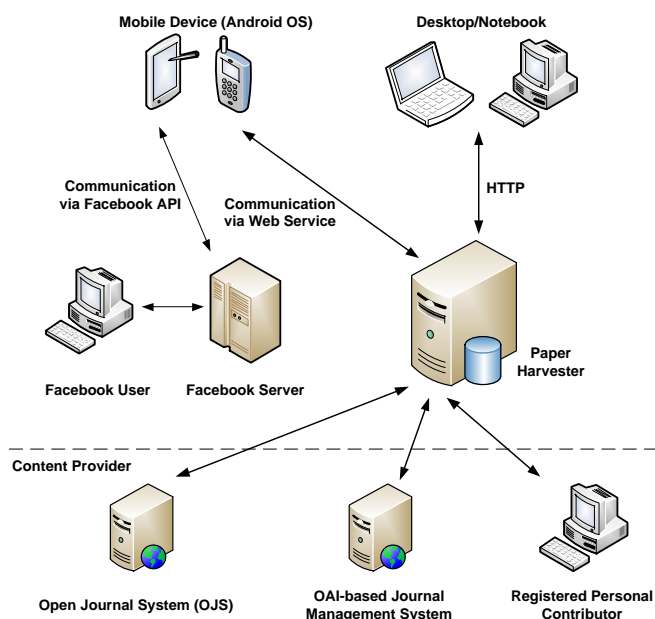


FIGURE I. ISCD PROJECT

The harvester part of the project is using Open Harvester System (OHS). OHS is a metadata harvester and indexing system developed by Public Knowledge Project. OHS implements Protocol for Metadata Harvesting from Open

Archive Initiative (OAI-PMH) [3]. By using OAI-PMH, OHS collects paper and journal’s metadata from journal database that also implement OAI-PMH. Originally OHS does not support citation database, so this research try to improve OHS by adding capability to store citations and link them to build citation network.

II. OPEN ARCHIVE INITIATIVE – PROTOCOL FOR METADATA HARVESTING

OAI-PMH is basically an implementation of REST (Representational State Transfer) based Web services protocols. REST architecture consists of a server and a client. REST client in the OAI-PMH uses GET and POST operations to retrieve metadata collections that are stored by the REST server. Data is sent from the server to the client in the form of XML documents as shown in Figure 2.

OAI-PMH uses verbs to identify the type of operation requested by the client to the server [4]. Verbs are used to determine the metadata formats that are supported by the repository, to fetch paper metadata from the server, or to know the categories provided by the repository server. Complete verb list is shown in Table 1.

TABLE I. REQUEST VERB LIST USED IN OAI-PMH

Verb	Function
GetRecord	Retrieve one metadata record from the server
Identify	Getting the OAI-PMH protocol versions supported by the server, the email administrator, record removal system, and the level of date detail
ListIdentifiers	Retrieve a list of papers’ header.
ListMetadataFormats	Get metadata format supported by server
ListRecords	Retrieve papers’ metadata based on date or set criteria
ListSets	Get paper’s set (category)

Although PKP-OHS support all OAI-PMH metadata elements, the harvester system will only use title, creator, and journal name including journal number or edition metadata elements from harvested XML. Additional information about paper (bibliographic and citations metadata) are provided by the parser used by the harvester.

```

<OAI-PMH
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<request identifier="oai: 10.1.1.40.5588"
metadataPrefix="oai_dc"
verb="GetRecord">oai2</request>
<GetRecord>
<record>
<header>
<identifier>10.1.1.40.5588</identifier>
<datestamp>2009-04-11</datestamp>
</header>
<metadata>
<oai_dc:dc
xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>A Method for Obtaining Digital Signatures and Public-Key Cryptosystems</dc:title>
<dc:creator>R.L. Rivest</dc:creator>
<dc:creator>A. Shamir</dc:creator>
<dc:subject>the difficulty of factoring the published divisor</dc:subject>
<dc:description>An encryption method is presented ...</dc:description>
<dc:date>2009-04-11</dc:date>
<dc:format>application/postscript</dc:format>
<dc:type>text</dc:type><dc:identifier>http://cite.seerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.5588</dc:identifier>
<dc:source>http://www.matha.mathematik.uni-dortmund.de/~fv/diplom_i/ars78.ps</dc:source>
<dc:language>en</dc:language>
<dc:relation>10.1.1.116.2833</dc:relation>
<dc:relation>10.1.1.115.3569</dc:relation>
<dc:rights>Metadata may be used without restrictions as long as the oai identifier remains attached to it.</dc:rights>
</oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

FIGURE II. XML DOCUMENT AS SERVER’S RESPOND FOR CLIENT’S GETRECORD REQUEST

III. HARVESTER SYSTEM

A. Harvesting Process

Harvester system works by extending PKP-OHS to support citation processing and storing. Harvester system has three steps in collecting, parsing and storing papers’ metadata (Figure 3).

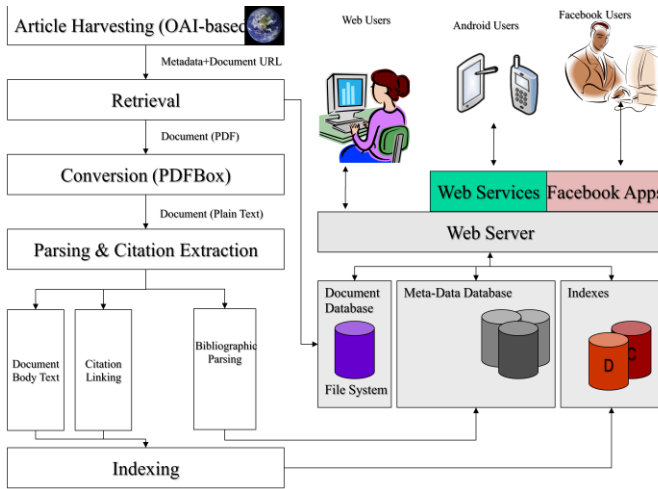


FIGURE III. HARVESTER & CITATION SYSTEM

1. Metadata Extraction

Harvester first has to have a list of content providers which are Indonesian online journal databases that support OAI-PMH. By using PKP-OHS, harvester retrieves bibliographic metadata of papers from content providers. As mentioned in chapter 2 harvester will only store title, creator, and journal name from harvested metadata into database.

This step also try to grab PDF files location for each paper from html files acquired from content providers' web server. Harvester uses regular expression to locate PDF files' URL address.

2. Citation Extraction

By using PDF files location from step 1, this step download the files and parse them to extract bibliographic and citation metadata. The harvester system utilizes other part of research project which is a parser that extract citations from paper's PDF files. The parser is an enhanced ParsCit system [5] that able to identify Indonesian language based bibliographic metadata and citations from papers [6]. For bibliographic metadata, the parser will provide author's name, affiliation, email, paper's title, abstract, and keywords. For citation the paper can provide paper's title from citations, and also the name of all authors, journal name and edition. How the parser works is explained in author's other project of [2] and [6].

3. Citation Citagraph

This step try to match citation metadata records with original paper records already stored in database. If there is a match the link is stored in table puslit_citagraph_citations. This step also try to find if the paper is self cited, that means that the authors cite their own paper. If self cited then field "self" will be given a value of 1.

B. Citation Information Output

To display citation information gathered by harvesting process, PKP-OHS has been modified. There are several modifications to PKP-OHS that can be accessed at <http://www.gudangpaper.net>:

1. Main page is able to display 10 most cited paper, and also link to see paper in more detail as can be seen at Figure 4.



FIGURE IV. TEN MOST CITED PAPERS

2. Search result page can show number of citations to the paper and also name and link of other papers that cite the paper as can be seen at Figure 5.

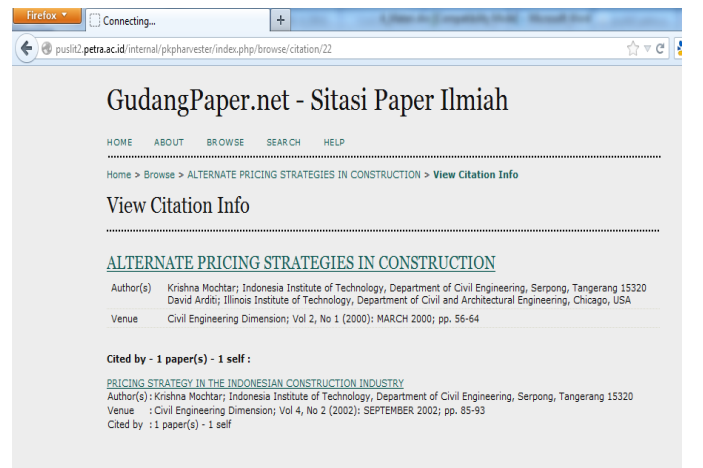


FIGURE V. LIST OF PAPERS THAT CITES DISPLAYED PAPER

3. Record Details page can show number of paper that cites, number of self-citation, a link to download the paper, and list of paper's citations as can be seen at Figure 6.

PENGARUH ELEKTROKINETIK TERHADAP DAYA DUKUNG PONDASI TIANG DI LEMPUNG MARINA

Cited by: 2 paper(s) - 1 self

Civil Engineering Dimension
[VIEW ARCHIVE INFO](#)

FIELD	VALUE
Authentication Code	dc
Title Statement	PENGARUH ELEKTROKINETIK TERHADAP DAYA DUKUNG PONDASI TIANG DI LEMPUNG MARINA
Added Entry - Uncontrolled Name	Daniel Tiandji ; Faculty of Civil Engineering and Planning, Petra Christian University Parvita Sri Wulandari ; Faculty of Civil Engineering and Planning, Petra Christian University
Uncontrolled Index Term Summary, etc.	elektrokinetika, pile foundation, bearing capacity, un-drained shear strength. Elektrokinetika is one of soft ground improvement methods to improve its bearing capacity. The objective of this research is to investigate the increase of friction and end bearing resistance of an embedded instrumented model pile in marine clay after elektrokinetika treatment. The initial geotechnical properties of soil surrounding the pile foundation is obtained by several soil laboratory tests. For pile bearing capacity analysis, friction and end bearing resistance of pile was measured by load cell measuring device, while un-drained shear strength analysis was measured by vane shear test. The result of this research showed that after 24 hours elektrokinetika treatment, the un-drained shear strength increased closer to the pile, and the bearing capacity of pile increased 14 times. Abstract in Bahasa Indonesia : Elektrokinetik adalah salah satu metode perbaikan tanah lunak yang diaplikasikan untuk meningkatkan daya dukung tanah lunak. Tujuan dari penelitian ini adalah untuk menyelidiki peningkatan tahanan friksi dan ujung suatu model pondasi tiang di lempung marina setelah dilakukan proses elektrokinetik. Karakteristik tanah pada kondisi mula-mula di sekitar pondasi tiang didapatkan dari beberapa pengujian di laboratorium. Untuk analisa daya dukung tiang, tahanan friksi dan ujung dari pondasi tiang diukur dengan alat pengukur load cell, sedangkan analisa kuat geser undrained diukur dengan pengujian baling-baling. Hasil penelitian menunjukkan bahwa setelah pada tanah dilakukan proses elektrokinetik selama 24 jam, daya dukung dari pondasi tiang meningkat 14 kali dan semakin dekat dengan tiang, kuat geser undrained juga semakin meningkat.
Publication, Distribution, Etc.	2006-02-23
Electronic Location and Access	application/pdf http://puulit2.petra.ac.id/ejournal/index.php/civ/article/view/16375
Data Source Entry	Civil Engineering Dimension; Vol 8, No 1 (2006); MARCH 2006; pp. 15-19
Language Note	en
Terms Governing Use and Reproduction Note	Copyright Civil Engineering Dimension, Journal of Civil Engineering Science Application
Download File	111.pdf

Citations :

- Shang, J.Q. and Masterson, K.L., An Electrokinetic Testing Apparatus for Undisturb/ Remolded Soils under In-Situ Stress Condition, Geotechnical Testing Journal, GTJODJ, Vol. 23, No. 2, June 2000, pp. 221-230
- Misic, S., Shang, J.Q., and Lo, K.L., Load-carrying Capacity Enhancement of Skirted Foundation Element by Electrokinetics, International Journal of Offshore and Polar Engineering, Vol. 13, No. 3, September 2003
- Richard, E.S. and Krishna, R.R., Effect of pH Control at The Anode For The Electrokinetic Removal of Phenanthrene From Kaolin Soil, Elsevier Science Ltd, December 2002, pp. 273-
- Shang, J.Q. and Lo, K.Y., Electrokinetic Dewatering of a Phosphate Clay, Journal of Hazardous Materials, 1997, pp. 117-133

FIGURE VI. RECORD DETAILS PAGE

C. Mobile Application Service

Users can also access the citation information from their Android based mobile device. Mobile application communicates with PHP web service (server) through HTTP protocol for requesting the database data.

```
{
  "products": [
    {
      "article_id": "111",
      "title": "PENGARUH ELEKTROKINETIK TERHADAP DAYA DUKUNG PONDASI TIANG DI LEMPUNG MARINA"
    },
    {
      "article_id": "120",
      "title": "THE PERFORMANCE OF TRADITIONAL CONTRACT PROCUREMENT ON HOUSING PROJECTS IN NIGERIA"
    },
    {
      "article_id": "125",
      "title": "Exploring Public Perception of Paratransit Service Using Binomial Logistic Regression"
    },
    {
      "article_id": "195",
      "title": "Seismic Properties of Moment-resisting Timber Joints with a Combination of Bolts and Nails"
    },
    {
      "article_id": "260",
      "title": "Effect of Staurosporine on the Intracellular Localization of Hepatitis B Virus Core Protein"
    },
    {
      "article_id": "291",
      "title": "Expression and Intracellular Localization Study of Wild Type HBV Core Protein and its Mutants Which Block Nucleocapsid Envelopment in HuH-7 Cells"
    },
    {
      "article_id": "495",
      "title": "Preparation Methods and Applications of CuO-CeO2 Catalysts: A Short Review"
    },
    {
      "article_id": "544",
      "title": "Design of a Compact and Versatile Bench Scale Tubular Reactor"
    },
    {
      "article_id": "22",
      "title": "ALTERNATE PRICING STRATEGIES IN CONSTRUCTION"
    },
    {
      "article_id": "57",
      "title": "PRICING STRATEGY IN THE INDONESIAN CONSTRUCTION INDUSTRY"
    }
  ],
  "success": 1
}
```

FIGURE VII. EXAMPLE OF JAVASCRIPT OBJECT NOTATION (JSON) FORMAT

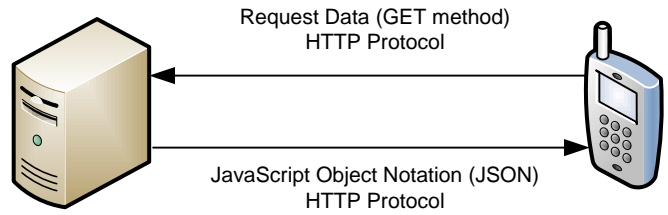


FIGURE VIII. DATA FLOW BETWEEN WEB SERVICE (SERVER) AND MOBILE DEVICE

Server gives response with deliver the JavaScript Object Notation (JSON) containing the requested data (Figure 7). Then mobile application extract the JSON and get the information to be displayed to users. (Figure 8)

Some features on mobile application are :

- View/browse the journal or conference data. This section provide to users the lists of collection of journal or conference papers.

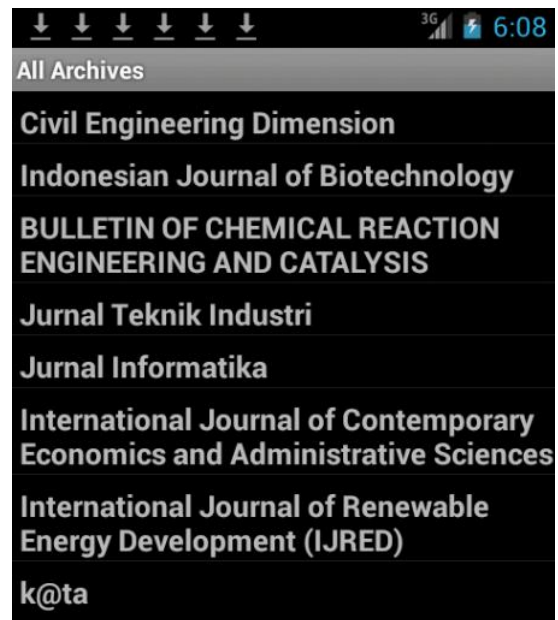


FIGURE IX. LIST OF ARCHIVES

- Search paper. Users can search the paper collection with some related keywords.

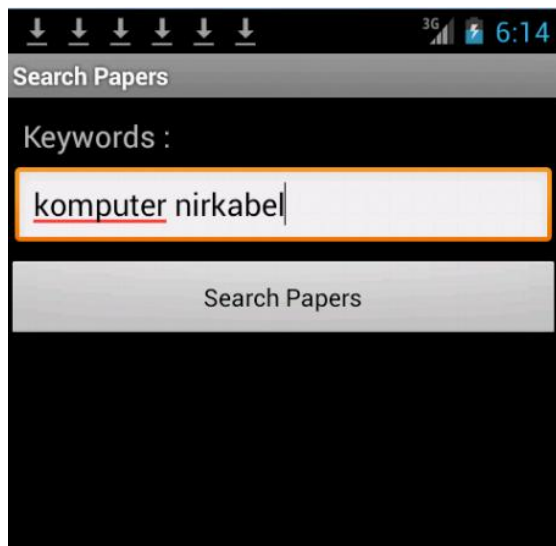


FIGURE X. FEATURE OF SEARCHING PAPER

3. View top cited paper. Users can view the top 10 most cited paper from Gudang Paper collection.



FIGURE XI. LIST OF TOP CITED PAPERS

4. Share citation data to Facebook. This feature enable users to share the citation data (title, abstract, venue, and link of a paper) to their Facebook's wall.

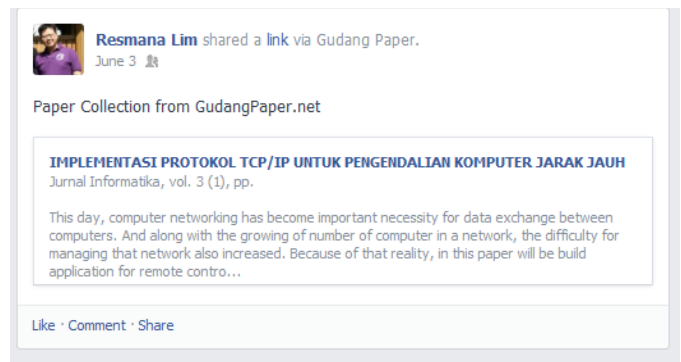


FIGURE XII. LINK SHARING RESULT ON FACEBOOK PAGE

5. Share citation data to e-mail, messenger, and other social networks.
6. Fulltext download and link to original site. Users can download full text paper (PDF file) and browse the original site.



FIGURE XIII. DETAIL OF PAPER WITH DOWNLOAD FULL PAPER AND VIEW FULL SITE BUTTON

IV. CONCLUSION

This research is able to extend the capability of PKP-OHS to process, store, and provide information about paper citations. Extended PKP-OHS needs parser that able to identify Indonesian language based citation and bibliographic metadata from papers. This research provides paper citation access via Web and Android application for mobile user.

REFERENCES

- [1] Henning, Victor., Reichelt, Jan., "Mendeley-A Last. fm For Research?". In: IEEE Fourth International Conference on eScience. IEEE, 2008. p. 327-328.

- [2] Lim, Resmana., Wibowo, Adi., Sutjiadi, Raymond., Eko Oktian, Yustus., "Indonesian Citation Based Harvester System", In: Makassar International Conference on Electrical Engineering Informatics (MICEEI 2012).
- [3] Owen, GW Brian., and Kevin Stranack., "The Public Knowledge Project and the Simon Fraser University Library: a partnership in open source and open access." *The Serials Librarian* 55.1-2, 2008. p. 140-167.
- [4] Lagoze, Carl., and Herbert Van de Sompel., "The making of the open archives initiative protocol for metadata harvesting." *Library hi tech* 21.2 (2003): 118-128.
- [5] Councill, Isaac G., Giles, C. Lee., Kan, Min-Yen., "ParsCit: An open-source CRF reference string parsing package." In: *Proceedings of LREC*. Vol. 2008. European Language Resources Association (ELRA), 2008.
- [6] Lim, Resmana., Wibowo, Adi., Sutjiadi, Raymond., Eko Oktian, Yustus., "Pengembangan Paper Citation Extraction Bahasa Indonesia Berbasis Parscit", In: Seminar Nasional Teknologi Informasi (SNTI 2012).