



# EVALUATION OF ARTICLES PUBLISHED IN MENDELEY AND CROSSREF IN RELATION TO THE GOOGLE SCHOLAR PAGES

Adian Fatchur Rochim and Riri Fitri Sari

Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Kampus Depok, Indonesia

Email: [adian.fatchur@ui.ac.id](mailto:adian.fatchur@ui.ac.id)

## ABSTRACT

This paper aims to show the performance of a researcher from their published articles. Our software crawled 10 (ten) most cited articles on the Google Scholar (GS), Mendeley and CrossRef with several of crawling methods. The method used in data retrieval is scrapping due to the limitations on the Application Programming Interface (API) provided by the Google search engine. To retrieve the Digital Object Identifier (DOI) data from CrossRef, the API method has been used. In order to count the number of reader of paper on the Mendeley we used the API method. We used the R programming language, Python and Bash scripting shell. The operating system was based on Ubuntu 8.04 Linux and Mac OS. The Apache webserver were used to serve the website and we used the MySQL database to store the data. The database of MySQL is used for interfacing between R with the PHP language purposes. The Hypertext Preprocessor (PHP) is used for server-side scripting. Data was obtained by scrapping the best 10 articles from 100 Indonesia's scientists indexed on the GS. Firstly, the data samples (S') were obtained from the list of Indonesian scientists in Webometrics as the input of the GS scrapping. Secondly, the data resulted (S'') were used as the input of the Crossref's API query to obtain the DOI of each article. Finally, the DOIs were used as the input for the API query to get the number of the result to show the number of readers of each to article on Mendeley. The software produced can crawl the data from Google Scholar, Crossref and Mendeley reader count.

**Keywords:** mendeley, reader count, google scholar and crossref.

## 1. INTRODUCTION

Online scientific publications are increasingly used by higher education institutions and research institutions. The Result of scientific papers produced can be accessed freely online at no charge by the public [1].

This research movement with the Open Access Initiative (OAI), which allows people to online access to scientific production that contain internal research results, journal paper, conference proceeding, and chapters of a book or thesis [2]. The impact factor was measured by how many articles cited/referenced by others and written back into the published article [3]. The article cited by others who work as academics or researchers will add citation value of the article. The increase number of other researchers that refer to that paper, to cite the paper and to produce new published article would increase the impact of that cited article [4].

For example, an article may be of an interest for nurse and this support their professional conduct a nurse. An article will be helpful because it is used, and then they are shared to their peers, discussed via the Internet through blogs, website and through social media forum [5]. This means that the article has a good impact factor. Citation performed by people who does not work as researchers/academics do not add value to the impact of a cited article [6]. Scientometrics 2.0 / Altmetrics is a quantitative study to measure the results of research, scientific communication, also measure the finding that has been discussed in the social media [7].

Networking and collaboration technology based on xml-based applications makes more research activities to grow rapidly. The initiation of the Open Archive Initiative accelerated the growth of the number of research publication in scientific journals. It also facilitated the tracking of the citation [8]. A sample of an application that is widely used to write scientific publications is Mendeley.

Mendeley is a reference manager that can be used to organize the result of the reference citations in an article [9]. Mendeley is similar to applications such as Zotero or EndNote. Features of Mendeley different from the others as it has a service feature interaction with other users. The service communication interactions with other users are similar to the feature on the social media application. Therefore Mendeley is called as a social media network research [10].

Mendeley is the reference manager, which can be used to manage the references of a paper. The difference of the Mendeley to other research tools such as Zotero and End Note is that Mendeley used "social media" function to create the network of authors. Mendeley is a the popular reference manager to manage of the references to produce scholarly articles [9] [11]. Open Archive Initiative develops and promotes to improve the access to e-print archives as a means of increasing the availability of scholarly communication [8].

Our motivation in doing this experiment is based on the manifesto issued by Jason Priem called "Altmetrics Manifesto" [12]. Altmetrics Manifesto tried to find data on the Internet using a technique called scientometrics/webometrics 2.0. The purpose of the altmetrics is to raise the impact the citation of an article. The paper is not only used to produce scholarly product but also improve the professional approach to work. For example a nurse who read an articles on Mendeley but she/he does not write paper and only uses the Mendeley in their work [12].

The following is some information about the research profile obtained from a search engine such as Google Scholar (GS) and Microsoft Academic Search (MAS). Our research focused on the GS search engine because, it is free to any users and accessible to anyone, it has features such as search for journal articles, search



libraries, set up alerts and get quick citations etc. Furthermore, most people used it. Information about the profile of researchers in both search engines shows the number of articles indexed in the search engine. Google Scholar produces the article title, author and co-author, as well as the citation obtained such as a h-index and h-10 index of the researcher.



Figure-1. An example of author profile from google scholar.

In addition to the Google Scholar, is another search engines shown below is Microsoft Academic Search. MAS also provided users with the facility to see of the researcher's profile. The data is displayed on the Microsoft Academic search is added with information of articles, information of the researchers, and also the network of the researchers with co-authors.

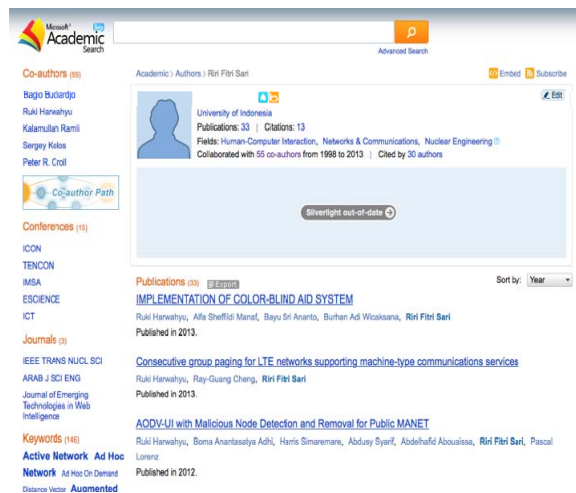


Figure-2. An example of the author profile from Microsoft academic search.

Information obtained from the MAS networks and relationships with co author on the menu of the MAS co-author graph can be seen in Figure-3.

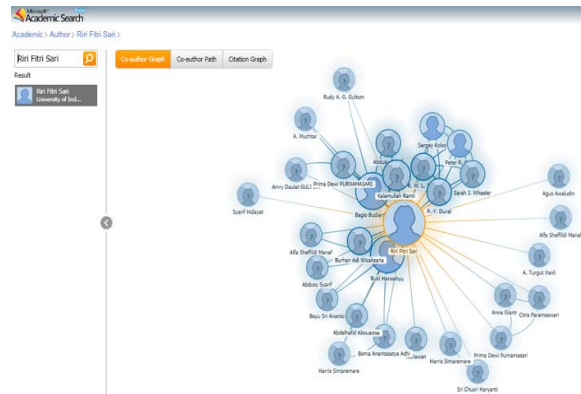


Figure-3. Networking of author and co-author on the MAS.

In this work use Google Scholar, Mendeley reference manager, and the DOI CrossRef indexing. Table-1 shows academic search engines application.

Table-1. The search engines.

Company/ (Name Search Engine)	Academic Search
Microsoft Corp. (Bing.com)	Microsoft Academic Search (MAS) <a href="http://academic.research.microsoft.com">http://academic.research.microsoft.com</a>
Google Corp. (Google.com)	Google Scholar <a href="http://scholar.google.com">http://scholar.google.com</a>

## 2. BIBLIOMETRICS, SCIENTOMETRICS AND WEBOMETRICS

### a) Bibliometrics

Bibliometrics was made possible by Garfield, and becoming established as a field in 1958 [13]. Bibliometrics is a statistical analysis of the bibliographic data, with a focus on citation analysis of a research/publication. The impact of study conducted by researchers, group or institution. The output of the measurement of the statistical analysis in bibliometrics is the number of publications/research cited in [14]. Bibliometrics includes a technique widely used in analyzing the frequency of use of words, citation analysis, co-author citations and documents as a simple counting of the number of publications made by researchers, research groups and countries [13].

Table-2 displays information the data from the bibliometric tool that can be used to collect / seek data on the bibliometric.

**Table-2.** Tools to access the bibliometrics data[14].

Tool	Feature	Metrics
SciVal	SciVal is a subscription based research performance assessment tool which uses data from Scopus.	<ul style="list-style-type: none"> <li>. Citation Count □</li> <li>. Citations per Publication □</li> <li>. H-index □</li> <li>. Journal Count □□</li> <li>. Number of Citing Countries</li> <li>. Collaboration Impact</li> <li>. Academic-Corporate</li> <li>. Collaboration Impact</li> </ul>
Scopus	Scopus is a subscription based citation database of peer reviewed literature. only available for papers published from 1996	<ul style="list-style-type: none"> <li>. Scholarly output</li> <li>. Citation counts</li> <li>. H-index</li> <li>. Number of co-authors</li> <li>. SC Imago Journal Rank</li> <li>. Scopus SNIP</li> </ul>
Tool	Feature	Metrics
	onwards.	
Web of Science	Web of Science is a subscription based citation database journals and conference proceedings.	<ul style="list-style-type: none"> <li>. Scholarly output</li> <li>. Citation counts</li> <li>. H-index</li> <li>. Journal impact factor</li> </ul>
Journal Citation Reports	Journal Citation Reports is a subscription based resource which allows to evaluate and compare journals using citation data	<ul style="list-style-type: none"> <li>. Journal impact factor</li> <li>. Most frequently cited journals in a field</li> <li>. Highest impact journals in a field</li> <li>. Largest journals in a field</li> </ul>

**b) Scientometrics**

In 1969, Vassily V. Nalimov and ZM Mulchenko Russian researchers coined the word "Scientometrics" ("Naukometriya") [15]. The terminology focused on the study of literature related to all aspects of science and technology. In 1978 The Journal Scientometrics is published by Tibor Braun from Hungary [15]. In 2010 Jason Priem wrote the Scientometrics 2.0, as a combination of scientometrics and web 2.0 [5].

**c) Webometrics**

Webometrics is a study of quantitative analysis of the data available on the Internet in the form of hyperlinks, and web data structure. Cybermetrics or Webometrics is a

metric of study on the Internet technology/electronic communications. The study phase of quantization metric data on the Internet is still at the level of discussion on this validity adaptation of traditional and electronic quote. The focus of the discussion is on the impact factor to calculate the problem of the web and the impact validity measurement of the web [16].

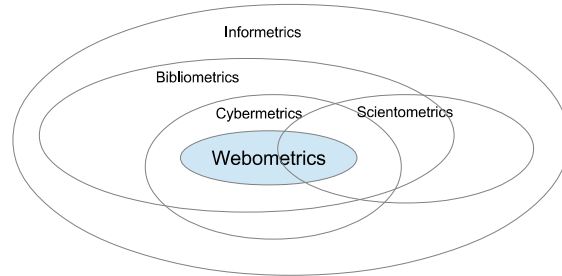
**Figure-4.** Webometrics and cybermetrics in the context of information science [17].

Table-3 that contains the applications of the website used by the researchers to write their scientific work and communicate with others [5]

**d) H-index**

In 2005, J.E. Hirsch wrote his work on the quantitative measure of the result of research investigation using H-Index terminology. H is from Hirsch (2005). H-index is an index that measures the quantity of researchers in generating scientific work. Some parameters of the h - index is as follows [4]:

- The number of paper/article ( $N_p$ ), to measure its productivity.
- The number of citations obtained from the resulting paper ( $NC, tot$ ) to measure the impact value divided by the total impact.
- The number of citations per article (ratio of  $N_c, tot$  to  $N_p$ )
- Articles significances, defined as the number of articles that have the larger number of citations  $y$  (i.e. for example,  $y = 50$ ). An example is the h-10 index released by Google Scholar.
- The number of citations is from the number of the most cited article  $q$ .

The total number of citations after  $n+1$  year is defined by the following parameter [4]:

$$N_{c,tot} = \sum_{j=1}^n p c_j = \frac{p c n(n+1)}{2} \quad (1)$$

**Table-3.** Web 2.0 dan Scientometrics [5].

Application name	Application	Academic application
Social bookmark	Delicious	Cite-ulike
Social collection management	iTunes	Mendeley Zotero
Social news	Friendfeed	Faculty of 1000
Publisher-hosted comemnt spaces	Most web 2.0 applications	British medical journal Plos Biomed Central
Microblogging	Twitter	
User-edited reference	Wikipedia	Encyclopedia of Life Scholarpedia Citizendium
Blogs	Wordpress	Research blogging
Social networks	Facebook Myspace	Nature Networks VIVOweb
Social video	Youtube Vimeo	SciVee

**e) Google scholar**

Google Scholar is the service from the Google corporation to provide us with search result of the scholarly literature from different disciplines and widely sources such as articles, theses, books, abstracts and court opinions. The articles from the academic publishers, professional societies, online repositories, universities and institutional web sites [18].

**f) Crossref**

Crossref is a platform from the collaboration of the linking references that function as a short of digital linking to the articles. The complete name is Crossref Digital Object Identifiers (Crossref DOI). The mission of the Crossref is "be a trusted collaborative organization with broad community connections; authoritative and innovative in support of a persistent, and sustainable infrastructure for scholarly communication" [19]. We gather the data of the DOI from resolver Crossref, because it has also tracking citations feature.

The DOI is a unique code to identify an article. If an article change the domain name, still identified by the DOI. DOI system guarantee that the digital copies of articles. This is the reason, why we used the DOI to identify the articles.

**g) Mendeley**

Mendeley is a reference manager. Mendeley is a dynamic global research collaboration platform that academics to improve their workflow and accelerate the pace of scientific discovery. We can download and Mendeley to interact with the papers from the other users, to follow and to communicate using social media. The Mendeley lead us to measure how many time the articles have been read by others [20]. Mendeley is a trademark of Elsevier limited, our research did not support by

Mendeley. Our research is independent and used Mendeley as a sample of the academic social media usage.

**3. METHODOLOGY**

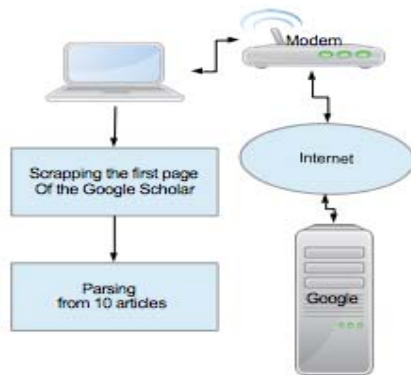
The following is used to measure the number of the articles have been read in Mendeley, Google Scholar and Crossref. Stages taken in conducting the experiment is as follows:

- Collecting and making a list of names of researchers that obtained sampling from a list of the 100 Indonesian researchers in Webometrics ranks of scientists (<http://webometrics.info>), to get a list of names of researchers [21]. We pruned the data that we got from Webometrics, because my assumption: the top 100 of the Indonesia researchers sorted by webometrics as a good sample for the Indonesia researchers population.
- From data obtained we used the article's title to scrap the search engine using Google Scholar with the input data of the name of researcher. The application used is the Python programming language [22].
- Finding the data of DOI from CrossRef for identifying all articles obtained. The method used the API provided by CrossRef API. R language is used to create the script.
- Calculating the number reader of the article on the Mendeley API Query method using R language. The Mendeley API can be used for authenticating applications on Mendeley.
- Storing data and calculating the result of the number of articles on Google Scholar Citation and CrossRef as well as the number of readers of the article on Mendeley using MySQL database application.
- Displaying the results in a bar chart representation of the picture and the list of researchers while using the programming language PHP and Apache webserver applications.

The Method of the Google Scholar scrapping is by taking the first pages found in Google Scholar that the researcher's name as a keyword. The data obtained is parsed as a selected data types to be used. Figure-5 shows the scrapping method using scholar.py [22]. Christian Kreibich, 2013, writes Scholar.py. It is a python module that includes an inquirer and parser for Google Scholar's search results, it as an open source code. The title, URL, year and number of citations can be extracted and saved as text file [22].

Scholar.py parsed the result from the application. Then, we selected the data of the article titles and the number of citations using the Bash shell scripting using awk. Awk is simple applications with an input such as the C language that we can use to process the word [23].

The output of this process produces a list of article titles, which are then used for further processing the input data name using API to query the CrossRef.



**Figure-5.** Scraping method using scholar.py by Christian Kreibich, 2013 [22].

Publication title, main online URL, number of online version can be extracted using scholar.py. Figure-5 shows block diagram of the extractor GS data. "Scholar.py" parses only the first search result page. Result of the scholar.py process, we filter as shown at algorithm 1, eliminate several characters, and then stored. □

**Algorithm 1.** Algorithm the process of scrapping of the data from Google Scholar

1. **Input** an author name
2. Read author name
3. Run scrip scholar.py with variable name
4. Eliminate char " ", ":", and " , "
5. Store the result

A query using the API of Crossref is provided using the R language and Rcrossref package [24]. The usage of the Rcrossref can be seen at algorithm 2:

**Algorithm 2.** Query data to Crossref

1. **Input** -> data article title
2. **Process**, crawl
3. **Parsing** data result from crawling process, number of the DOI
4. **Output** -> store DOI number to data

Parsing data at Algorithm 2, Parsing from a crawl result data file, we removed several data fields and the number of the DOI only to store.

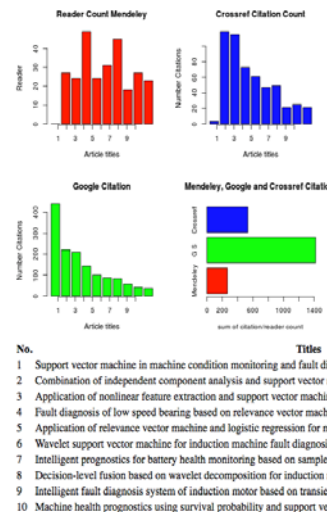
**Algorithm 3.** Crawl data from Mendeley

1. **Input** (token, DOI)
2. **Process**, crawl readers from Mendeley
3. **Process**, export raw to character
4. **Store** data from Mendeley
5. **Output**, data table (DOI, article title, Mendeley Reader)

#### 4. RESULTS AND DISCUSSION

The results obtained are a list of DOI numbers for each article. Articles that do not have a DOI cannot be processed. Query using the API on the features that have been provided Mendeley using language R. The following script is our algorithm for the API to Mendeley to get the number of the article readers.

Finally, the results of the three processed the query is stored in a MySQL database and displayed via the localhost website.



**Figure-6.** The number of readers in Mendeley, Google scholar, and Crossref is as the result of the process.

Figure-6 is shown the citation number data from GS (Google Citation), the number of the Mendeley reader (Reader Count Mendeley), the citation number data of the resolver Crossreff (Crossref Citation Count). Each of them, then we represented as the fourth bar graph (Mendeley, Google and Crossref Citation), that shows of the total researcher performance.

#### 5. CONCLUSIONS

The result of the research contributes to provide the assessment of articles citations. The data representation of the results of citations and the number of readers who obtained the researcher can be accessed through the website by entering the list of researchers (plural in the form of a file) or the name of the researcher (single).



Future research will include the parameters provided by Scopus citations features through the APIs provided. The data representation of each researcher can be displayed on the complete web, from a search engine, citations and citation on altmetrics, bibliometrics / scientometrics 2.0.

We expect that our application will motivate researchers to working and improve their research, as the measurement show of the performance of publications can be used.

## REFERENCES

- [1] N. Li and X. Zhao, "The study on constructing institutional repository of university," Proc. - 2012 4th Int. Conf. Multimed. Secur. MINES 2012, pp. 689–692, 2012.
- [2] I. Farida, J. H. Tjakraatmadja, A. Firman, and S. Basuki, "A conceptual model of Open Access Institutional Repository in Indonesia academic libraries," *Libr. Manag.*, vol. 36, no. 1/2, pp. 168–181, 2015.
- [3] R. K. Pan and S. Fortunato, "Author Impact Factor: tracking the dynamics of individual scientific impact," *Sci. Rep.*, vol. 4, p. 4880, 2014.
- [4] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc Natl Acad Sci U S A*, vol. 102, no. 46, pp. 16569–16572, 2005.
- [5] J. Priem and B. H. Hemminger, "Scientometrics 2.0: New metrics of scholarly impact on the social Web," *First Monday*, vol. 15, no. 7, 02-Jul-2010.
- [6] P. Mcfedries, "Measuring the impact of altmetrics [Technically Speaking]," *IEEE Spectr.*, vol. 49, no. 8, pp. 28–28, 2012.
- [7] J. Priem, H. a Piwowar, and B. M. Hemminger, "Altmetrics in the wild: Using social media to explore scholarly impact," arXiv12034745v1 csDL 20 Mar 2012, vol. 1203.4745, pp. 1–23, 2012.
- [8] "Open Archives Initiative." [Online]. Available: <https://www.openarchives.org/OAI/OAI-organization.php>. [Accessed: 22-Dec-2015].
- [9] X. Li, M. Thelwall, and D. Giustini, "Validating online reference managers for scholarly impact measurement," pp. 461–471, 2012.
- [10] S. Haustein, V. Larivière, M. Thelwall, D. Amyot, and I. Peters, "Tweets vs. Mendeley readers: How do these two social media metrics differ?," *Inf. Technol.*, vol. 56, no. 5, pp. 207–215, 2014.
- [11] J. Priem, H. A. Piwowar, and B. H. Hemminger, "Altmetrics in the wild: An exploratory study of impact metrics based on social media."
- [12] J. Priem, D. Taraborelli, P. Groth, and C. Neylon, "Altmetrics: a manifesto," 2010.
- [13] M. Thelwall, "Bibliometrics to Webometrics," *J. Inf. Sci.*, vol. 34, no. 4, pp. 1–18, 2007.
- [14] C. Part, S. J. Rank, S. Snip, and J. C. Reports, "Bibliometrics : an overview," pp. 1–7, 2014.
- [15] W. W. Hood and C. S. Wilson, "The literature of bibliometrics, scientometrics, and informetrics," *Scientometrics*, vol. 52, no. 2, pp. 291–314, 2001.
- [16] J. M. Russell, S. Researcher, C. Universitario, and D. I. Bibliotecológicas, "Bibliometrics and Institutional Evaluation," pp. 1–20, 2002.
- [17] T. C. Almind and P. Ingwersen, "Informetric analyses on the world wide web: methodological approaches to 'webometrics,'" *Journal of Documentation*, vol. 53, no. 4, pp. 404–426, 1997.
- [18] "About Google Scholar." [Online]. Available: <https://scholar.google.co.id/intl/en/scholar/about.html>. [Accessed: 08-Dec-2015].
- [19] "crossref.org : : history/mission." [Online]. Available: <http://www.crossref.org/01company/02history.html>. [Accessed: 08-Dec-2015].
- [20] "Awards and Reviews | Mendeley." [Online]. Available: <https://www.mendeley.com/review/>. [Accessed: 08-Dec-2015].
- [21] "Asia/Pacifico | Ranking Web of Repositories." [Online]. Available: [http://repositories.webometrics.info/en/Asia\\_Pacifico](http://repositories.webometrics.info/en/Asia_Pacifico). [Accessed: 09-Sep-2015].
- [22] "scholar.py — A Parser for Google Scholar." [Online]. Available: <http://icir.org/christian/scholar.html>. [Accessed: 08-Dec-2015].
- [23] B. D. Dougherty, A. Robbins, and S. Edition, *Sed & Awk*, vol. 16, no. 7, 2014.
- [24] T. Hart, S. Chamberlain, and M. S. Chamberlain, "Package 'rcrossref.'" 2015.