Design data collection tool and weighting classification of authors in their scholar outputs based on percent- contribution-indicated (PCI) method

by Adian Fatchur Rochim

Design data collection tool and weighting classification of authors in their scholar outputs based on percent-contribution-indicated (PCI) method

Design data collection tool and weighting classification of authors in their scholar outputs based on percent-contribution-indicated (PCI) method

A F Rochim[1], T Nugraha[1], A P Widodo[2], D Eridani[1], K T Martono[1]

[1]Department of Computer Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, 50275, Indonesia
[2]Department of Computer Science, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, 50275, Indonesia

Corresponding author: adian@ce.undip.ac.id

Abstract. The advancement of information technology development makes it researchers easier to read papers shortly after they are published online. A paper will be cited, usually within one to two years after the paper is published. The number of citations of a researcher's papers is currently an indicator of his/her scholarly impact. The measurement widely used is the H-index. The calculation of the impact of the researcher using the H-index method does not calculate the weighted classification of the author's position on each paper. The main author or co-authors score have the same weighted citation weighted of an article cited. The weighting of the number of citations obtained by the author has begun to get attention by Scientometricians in recent years. This is done to differentiate the value of the impact obtained by the main author and co-author. Since 2015, the Government of Indonesia has awarded a number of awards to Indonesian researchers who have the greatest impact value. Assessment of the impact value of researchers in Indonesia uses the Sinta Index version 1 (SI-1) which is based on the researcher's H-index value in Scopus, GS, and WoS. Problems arise when a researcher with most of his papers as a co-author gets an award, while another researcher who has a position as the main author does not get the award, even though the number of papers and the number of citations obtained are relatively similar. In 2020, we found that Sinta Index was upgraded, namely Sinta Index version 2 (SI-2). There are some differences between versions 1 and 2. This paper will propose a mechanism for calculating the impact of researchers using the weighting of the investigator's position based on the percent-contribution-indicated (PCI) method. The data source used was the Scopus database. To get and classified the source data, we developed a tool using API. For the rank of the researcher, we used data from the Sinta Index website. Collection software design created using PHP and Mysql. From the research results, it was found that the correlation test by Pearson's Method of the H-index values of APW and Scopus is 0,44; APW vs GS (0,39) and GS vs Scopus is 0.95. This means that APW method has a weak correlation with the H-index values of GS and Scopus. APW is independent of the two h-index calculations.

1. Introduction

The number of citations and the number of papers was combines as H-index [1]. H-index is an index to figure the profile of the authors. Institutions use the H-index to rank the authors or researcher staff. H-index is a well-known index that is used by the database indexers such as Clarivate Analytics, Scopus, and Google Scholar.

From the literature, we found some weaknesses of the H-index, including 1) the productive and perfectionist researcher were not accommodated by the H-index [2], 2) self-citation was calculated, 3) the citation weight of the main researcher is considered equal to other researchers, and the frequency of citation in a paper has not been considered [2][3][4][5]. Many H-index improvement proposals have been made. This includes the proposal of Egghe in 2006, which accommodates the impact value of perfectionist researchers [6]. Bi Hui in 2007 proposed the impact evaluation of researchers based on the year of publication of papers [7]. A new method was proposed by Abramo in 2014 to measure the impact of researchers with a practical economic interpretation called the MNCS and FSS methods [8].

New indicators to measure the impact of researchers were needed for a better evaluation. Rochim, et al. in 2018 proposed the RA-index as an alternative indicator of fairer-based bibliometrics to measure the impact of researchers [9]. Glatzel, in 2016, stated that it is important to consider some methods and models to accommodate the needs [10].

Indonesia is a developing country that used the H-index to measure the impact of the Indonesian Researchers. The Indonesian Ministry of Research and Higher Education (Kemenristekdikti) has created a tool called Sinta-score to measure the impact of researchers based on the H-index. They ranked Indonesian researchers based on the Sinta-index (S-score) algorithm [11].

Sinta index used two database sources i.e., Google Scholar, Web of Science (Clarivate Analytics) and Scopus. Google Scholar and Scopus have a strong correlation, in which the H-index of authors in Google Scholar are in average 2.45 times of their H-index in Scopus [12]. Martin, in 2018 stated that there had been no recent evidence of differences between Google Scholar, Scopus, and Clarivate Analytics [13].

1.1. Sinta Index Method

In 2017, the Indonesian Government released a researcher performance indicator named Sinta index [14]. The Sinta Index calculation method version 2.0 is used to measure the impact of Indonesian researchers. Data sources are from the H-index calculation from two sources, i.e., Google Scholar and Scopus.

The following is the Sinta Index algorithm version 1 [11]:

$$\text{Sinta Index v.1} = ((A * 40) + (B * 15) + (C * 1) \, C + (D * 4) + (E * 4) + (F * 16)) / \text{Divisor} \qquad (1)$$

A: Number of article documents in Scopus   D: Number of citations on GS
B: Number of non-article documents in Scopus   E: H-index value on Google Scholar
C: Number of citations on Google Scholar   F: H-index value in Scopus
Divisor: 102 (Current divisor value used), 31 January 2017

Sinta index was used to measure the author level of Indonesian Researchers and journal level. To measure the impact of author, sinta index (SI) uses a combination of the H-index values of Google Scholar, Scopus, the number of paper in the Scopus and Google Scholar were weighted for each parameter to produce the value of Sinta Score or Sinta Index version 1 (SI-1) [11].

The problem was found on the difference in the measurement of index value between SI-1, Sinta Index version 2 (SI-2), and H-index of the Indonesian Authors. Several Indonesian authors have a higher H-index value in Google Scholar and Scopus, but when we measured with the (SI-1) and vice versa. The SI-1 does not take into account various publications and practices of joint writing in different fields [15]. In 2020, we saw that SI-2 released by Indonesia Ministry of Research and Technology. In website of the SI-2, we cannot find the SI-2 time released. The SI-2 algorithm little bit difference than the SI-1 algorithm.

Sinta Index version 2 (SI-2) assigns several different weights to each paper according to the level of journal quartile. Journals with quartile level 1 and 2 are given a weight of 40, quartile 3 35 and quartile 4 30. While the number of citations is given the following weights: citations from Scopus 4 indexed journals and citations from google indexed journals 0.5. Then the national journals are given weight according to the journal level from SI. SI makes a ranking of journals i.e. Journals S1 through S6. Journal

weighting S1 and S2: 25, journal S3 and S4: 20 and journals S5 and S6: 15. Table 1. illustrates SI-2 algorithm. For example, an author has 1 paper in Q2 journal, two papers in Q3 journals and etc.

Table 1. Sinta index version 2 algorithm [16]

Category	Weight	Value	Score (Weight * Value)
Paper in Journal Scopus indexed, Q1	40,0	0	0
Publikasi Scopus Q2Paper in Journal Scopus indexed,	40,0	1	40
Publikasi Scopus Q3	35,0	2	70
Paper in Journal Scopus indexed, Q4	30,0	0	0
Paper in Journal Scopus indexed, non quartile	30,0	1	30
Paper in Scopus Non Jurnal (Conferences, Books)	15,0	10	150
Number of citations in Scopus	4,0	15	60
Number of citations in GS	0,5	372 / 372	186
Papers in S1 accredited Journal	25,0	0	0
Papers in S2 accredited Journal	25,0	0	0
Papers in S3 accredited Journal	20,0	0	0
Papers in S4 accredited Journal	20,0	0	0
Papers in S5 accredited Journal	15,0	0	0
Papers in S6 accredited Journal	15,0	0	0
Total Sinta Score V2 Overall			536

The assumption of the S-2 algorithm multiplies the number of papers with journal weight i.e. 40 for journal in quartiles 1 and 2, etc. The assumption was used to adopt some of the policies of Indonesia Ministry of Education and Culture for the academic promotion of Indonesian lectures (Tim PAK). In 2018, we have identified that a small number of Indonesian researchers conducted some activity of "*citation circles*" to increase their H-index values [17]. "citation circle" is an activity in which someone cites the work of his friends, and will get a citation for the same way [18]. This is a part of the "black hat" technique. The technique is not accepted or illegal for academics. Cartels/citation circle can be defined as follows:

a. The activity of an author that act as also a reviewer for multiple papers at the same time and a joint-work among friends in a peer review ring to increase the record of papers and citation numbers [19].

b. The activity of an author cite his/her friend's papers, and at the same time, these friends also cite the author's papers [18].

In order to prevent the activity of "creating citation circle", we recommend the weighting mechanism for author position in their number of citations of their papers. The citation data is weighted before it is calculated by the H-index method. In this work, we propose a weighting mechanism of author position in a paper based on Percent-contribution-indicated (PCI) method. To show comparison of the proposed algorithm, we compare rank of top authors in SI (without weighted in author position) and rank of the top authors after weighted. The weighting position of authors in a paper is important to prevent "cartels/citation circle".

There are four sections in this paper. The first section describes the background and purpose of weighting mechanism of author position and previous research. The second section describes the methodology used. The third section describes the results and discussion. And finally, the conclusion is presented in the fourth section.

2. Method

2.1. Weighting mechanism using SDC, EQ, FLAE and PCI methods

Tscharntke in 2007 classified the weighting for each author in a publication text into four weighting methods groups. The four groups are: 1) Sequence-determining-credit (SDC), 2) Equal Contribution (EQ), 3) First author-emphasis (FLAE) and 4) Percent-contribution-indicated (PCI) [20]. The methods are as follows:

a. Sequence-determine-credit (SDC), the weighting method is based on the model for the first author to get half of the maximum credit.
b. Equal Contribution (EC), weighting is calculated the same for all authors with a minimum score of 5%. The author's order corresponds to the alphabet.
c. First-last-author-emphasis (FLAE), the first author receives the full credit. The second author gets half of the maximum credit, and the rest of authors receive the maximum credit divided by the number of authors minus two.
d. Percent-contribution-indicated (PCI), the credit distribution mechanism uses a combination of FLAE and EDC. The first author is given 60% of the maximum credit. The second author receives 20% of the maximum credit. The third author gets 10%, while the fourth and fifth author receive each 5% of the maximum credit.

This weighting mechanism is proposed to give an appreciation to the first author and the corresponding authors. The proposed method combination of the PCI and EC methods. Accommodates the policies of the Indonesian Government in granting credits for scientific publications (Tim PAK Dikti) and combine with the PCI method.

The method of the weighting mechanism is based on the combination of PCI and EC methods. For example, one paper has ten citations, and written by four authors i.e. main author (1), corresponding author (1) and other authors (2). The citation calculation obtained by each author is different and based on the following proportions as follows. The main author and correspondent get the maximum publication index value of 100% of the publication index value.

$$\text{author's publication index value} = ma \; x \; 100\,\% \qquad (1)$$
$$\text{others author's citation value} = (ma \; x \; 50\%)/n \qquad (2)$$

ma value = the number of total citations of a paper; $\quad n$ = total number of the others author.

Others author get a value of 50% of the maximum value divided by the number of other authors. We called the combination of the PCI, EC methods and accommodate regulation of Tim PAK from government as author position weighting (APW) method.
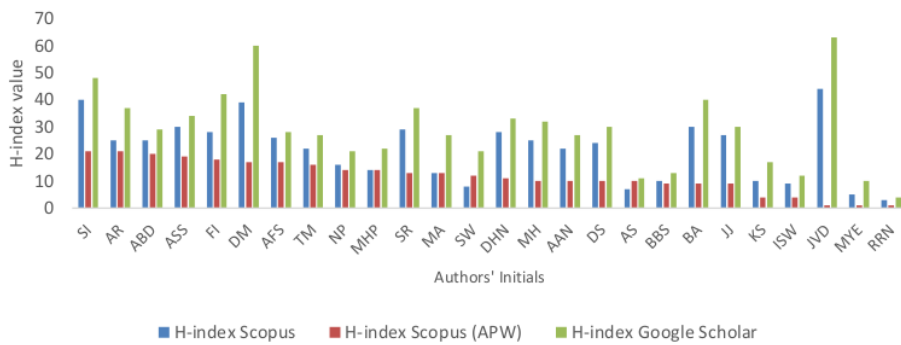

Figure 1. Distribution of H-index values based on sinta index rank.

The data source was obtained from Scopus using API Scopus. Design of the application to crawl and classify the author position by PHP and database backend using MySQL. Comparison of result of the APW method we used H-index Scopus, H-index GS and ranking of the top 26 Indonesian authors in Sinta Index website. We collected papers of the 26 top Indonesian authors of SI version. Data got at August 2020.

3. Result and Discussion

After we investigate data of the top authors in Sinta index, we can illustrate in figure 1 and 2. Figure 1. shows the rank of authors based on the Sinta Index Rank. And Figure 2. shows rank of authors based on the APW proposed method.

Figure 2. Distribution of H-index values based on H-index Scopus (APW)

In the figure 2. we can look that rank of author "JVD" went down from third to the rank of 25th. This ranking of the author was down because his/her position on the greatest number of papers are not as a main author. Furthermore, the most of papers have more than 200 authors in a paper, accordingly in the proposed method that the number of citations just only have half weighting and divided by all authors except the main author. We can use APW proposed method to measure author-level based on H-index, with an in-depth look at weighting authors' positions.

Figure 3. (a) and (b) shows weak correlation of H-index Scopus based on APW vs H-index Scopus and Google Scholar. the figures illustrate that H-index APW method, is slightly different view with others H-index Scopus and GS. Figure 3. (c) shows strong correlation of H-index Scopus and Google Scholar. The correlation is strong because data on the database of Scopus and Google Scholar are similar. However, GS has much more data than Scopus.
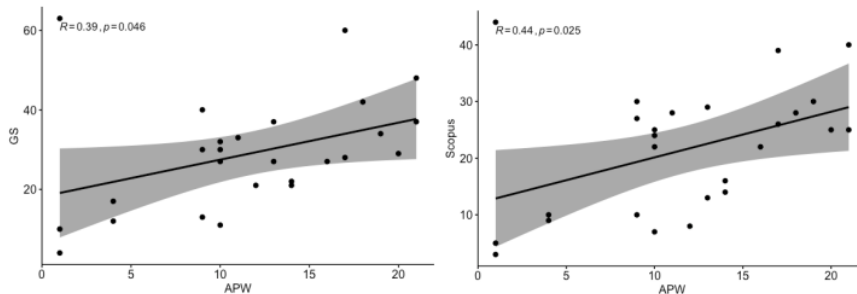
Figure 3.(a). correlation of H-index value of APW and GS, 3.(b). H-index values of APW and Scopus
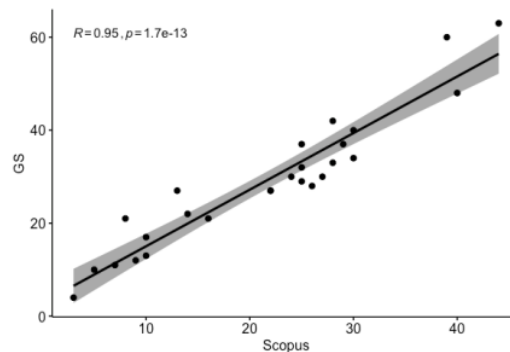
Figure 3.c. correlation of H-index of GS and Scopus

4. Conclusion

This work is a research in progress. The method was proposed to evaluate rank of authors based on position authors. The number of citations was filtered by the combination of PCI and EC. Result of correlation test by Pearson's Method of the H-index values of APW and Scopus is 0,44; APW vs GS (0,39) and and GS vs Scopus is 0.95. This mean that APW method has weak correlation with the H-index values of GS and Scopus. APW is independent of the two h-index calculations. The initial investigation shows that the APW method can be used to identify authors that is just a "piggyback", which have a higher ranking, high value of H-index Scopus and GS, but in fact, he/she does not have papers as a main author. In the future, it is planned to test the model by more Indonesian researchers data will be conducted.

References

[1]   Hirsch J E 2005 *Proc Natl Acad Sci U S A* **102** 46 16569–16572
[2]   Mesiar R and Gagolewski M 2016 *IEEE Trans. Fuzzy Syst.* **24** 6 1668-1672
[3]   Bai X *et al.* 2018 *PLoS One* **13** 3 1–17
[4]   Gagolewski M and Grzegorzewski P 2009 **81** 3 617–634
[5]   Zhu X, Turney P, Lemire D and Vellino A 2015 *Journal of Assoc. Inf. Sci. Technol.* **66** 2 408–427
[6]   Egghe L 2006 *Scientometrics* **69** 1 131–152
[7]   BiHui J 2007 *Chinese Science Bulletin.* **52** 6 855–863
[8]   Abramo G and D'Angelo C A 2016 *Journal of Informetrics* **10** 2 646–651
[9]   Rochim A F, Muis A and Sari R F 2018 *Desidoc Journal of Library and Information Technology* **38** 6 378–386
[10]   Glänzel W, Thijs B and Debackere K 2016 *Journal of Informetrics.* **10** 2 658–660
[11]   Lukman *et al.* 2018 *Science Editing.* **5** 2 135–141
[12]   Rochim A F, Muis A, Sari R F 2016 *Int. J. Educ. Pedagog. Sci.* **11** 10 1305–1311
[13]   Martín-martín A, Orduna-malea E and Thelwall M 2018 *Journal of Informetrics* **12** 4 1160–1177
[14]   Antara 2017 *Sinta-Index release* [Online]. Available: https://www.antaranews.com/berita/609670/menristek-luncurkan-sinta-portal-kinerja-peneliti [Accessed: 25-Sep-2018].
[15]   Chawla D S 2018 *Indonesia's scientists voice concerns about the country's researcher ranking system* 2018.

[Online].   Available: https://cen.acs.org/content/cen/articles/96/web/2018/12/Indonesias-scientists-voice-concerns-countrys.html. [Accessed: 07-Feb-2019].

[16] Ministry of Research of Indonesia, "Sinta Index version 2," *2020*. [Online]. Available: http://sinta.ristekbrin.go.id. [Accessed: 17-Sep-2020].

[17] Rochim A F and Sari R F 2019 in *17th International Conference on Scientometrics and Informetrics, ISSI* pp. 2608–2609

[18] Kienc W 2015 Should you care about your h-index, and if so, how to improve it? | Open Science," 2015. [Online]. Available: https://openscience.com/should-you-care-about-your-h-index-and-if-so-how-to-improve-it/. [Accessed: 11-Oct-2018]

[19] Gamboa C 2014 *SAGE statement on Journal of Vibration and Control Sage Publishing* [Online]. Available: https://uk.sagepub.com/en-gb/asi/press/sage-statement-on-journal-of-vibration-and-control. [Accessed: 06-Feb-2019]

[20] Tscharntke T, Hochberg M E Rand T A, Resh V H and Krauss J 2007 *PLoS Biol.* **5** 1 0013–0014

Design data collection tool and weighting classification of authors in their scholar outputs based on percent-contribution-indicated (PCI) method

PRIMARY SOURCES

1 Ziyun Cheng, Hao Li, Huihui Xu, Li Dong, Hongyuan Tao, Huichun Hua, Shenghua Wang. "Identification of market power abuse in China's electricity market", IOP Conference Series: Materials Science and Engineering, 2020
Publication
2%

2 Sunarno Sunarno, Purwanto Purwanto, Suryono Suryono. "Trend Analysis of NOX and SO2 Emissions in Indonesia from the Period of 1990 -2015 using Data Analysis Tool", Advances in Science, Technology and Engineering Systems Journal, 2020
Publication
1%

3 Richard S. J. Tol. "The h-index and its alternatives: An application to the 100 most prolific economists", Scientometrics, 04/16/2009
Publication
1%

4   He, B.. "Mining patterns of author orders in scientific publications", Journal of Informetrics, 201207

Publication

    1 %

5   Jian Xu, Ying Ding, Min Song, Tamy Chambers. "Author credit-assignment schemas: A comparison and analysis", Journal of the Association for Information Science and Technology, 2016

Publication

    <1 %

6   Nickalus R. Khan, Clinton J. Thompson, Douglas R. Taylor, Garrett T. Venable, R. Matthew Wham, L. Madison Michael, Paul Klimo. "An analysis of publication productivity for 1225 academic neurosurgeons and 99 departments in the United States", Journal of Neurosurgery, 2014

Publication

    <1 %

7   Leo Egghe. "Comparative study of h-index sequences", Scientometrics, 04/17/2009

Publication

    <1 %

8   Mauricio Lima Barreto, Erika Aragao, Luis Eugenio Portela Fernandes de Sousa, Taris Maria Santana, Rita Barradas Barata. "Diferencas entre as medidas do indice-h geradas em distintas fontes bibliograficas e engenho de busca", Revista de Saúde Pública, 2013

    <1 %

Publication

9 Janet Lee. "Use of the *h* index in neurosurgery", Journal of Neurosurgery, 04/24/2009
Publication

<1 %

10 Marcelo P. Dabós, Ernesto R. Gantman, Carlos J. Fernández Rodríguez. "The Prestige of Social Scientists in Spain and France: An Examination of Their h-Index Values Using Scopus and Google Scholar", Minerva, 2018
Publication

<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

Design data collection tool and weighting classification of authors in their scholar outputs based on percent-contribution-indicated (PCI) method

FINAL GRADE

/0

GENERAL COMMENTS

Instructor