A semantic interpretations of citation as assumptions

Mahyuddin K M Nasution¹, Marischa Elveny¹, Indra Aulia¹

¹Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Padang
Bulan 20155 USU, Medan, Indonesia
E-mail: mahyuddin@usu.ac.id

Abstract. Citation is specifically a problem in the world of publication after being used as a benchmark for ranking the writing quality, or indirectly ranking the author. This paper aims to reveal some of the assumptions interpreted from several data sets about citations, especially self-citations. Self-citation becomes a special problem when it becomes a measure for the performance of countries, journals, institutions, and authors. So far, there are links between assumptions about self-citation for the country, journal, and author.

1. Introduction
Every technology has a bad and good side, and this depends on the purpose of its use and functioning [1]. Likewise, it relates to citations as a tool used to determine the author’s eligibility based on its application of the H-index [2], or to see the continuity of the development of knowledge either in terms of research roadmap or on the downstream side [3].

Scientific publication is the basic evidence of the development of knowledge [4], which is directly as an outcome of research or not, which will indirectly lead to innovations needed in the development of a nation [5]. However, citations in scientific publications have the potential to be misused by writers or other people who mistakenly use or consider citation functions [6]. Therefore, this paper aims to explain the good and bad sides of citation semantically, especially self citation in particular, namely the advantages and disadvantages refer to the writing itself.

2. A review
Every scientific publication is designed based on the following principles: "The writing of scientific paper requires conceptual bridges from what is already known,..." [7, 8]. This explains that a scientific document is born from the author based on what is already known by the author so that the scientific work has a problem statement, objectives, and methodology [9]. The interests of a scientific work must be based on a review of the related literature so as to produce a problem statement. Meanwhile, the transformation of a problem statement into an objective in a scientific work is supported by reviews so that it presents a target. Whereas the methodology of delivering problem statements becomes results or conclusions, and the presence of methodologies requires review so that the model/approach/method is adjusted to the problem to be solved [10].

The review or more than it, state of the art, is things that are known, and what is known to be sourced from a bibliography or reference, which will specifically assess the experience of an author [11]. Although every scientific publication requires study, it is even more important that every scientific publication requires citation (to be cited and cited) [12].
| No. | Countries               | Doc     | Citable Doc | Citations | Self-Citations |
|-----|-------------------------|---------|-------------|-----------|----------------|
| 1   | United States           | 11,036,243 | 9,875,662   | 267,612,868 | 122,087,837    |
| 2   | China                   | 5,133,924  | 5,052,579   | 39,244,368  | 21,831,514     |
| 3   | United Kingdoms         | 3,150,874  | 2,705,067   | 68,803,194  | 15,755,046     |
| 4   | Germany                 | 2,790,169  | 2,590,028   | 54,834,760  | 13,548,169     |
| 5   | Japan                   | 2,539,441  | 2,437,565   | 39,049,963  | 10,407,744     |
| 6   | France                  | 1,967,157  | 1,837,639   | 37,865,266  | 8,085,273      |
| 7   | Canada                  | 1,594,391  | 1,446,619   | 34,945,309  | 6,216,383      |
| 8   | Italy                   | 1,583,746  | 1,451,214   | 28,548,485  | 6,597,300      |
| 9   | India                   | 1,472,192  | 1,379,217   | 12,637,866  | 4,329,674      |
| 10  | Spain                   | 1,256,556  | 1,156,724   | 20,661,273  | 4,705,368      |
| 11  | Australia               | 1,226,552  | 1,093,833   | 23,347,703  | 4,866,812      |
| 12  | South Korea             | 1,004,042  | 973,360     | 12,299,582  | 2,501,499      |
|     | **Sub-total**           | **34,755,287** | **31,999,507** | **639,850,637** | **220,932,619** |
| 32  | Singapore               | 265,452   | 246,176     | 4,786,877   | 557,083        |
| 34  | Malaysia                | 248,457   | 239,537     | 1,615,633   | 421,749        |
| 42  | Thailand                | 156,829   | 148,862     | 1,740,576   | 273,578        |
| 52  | Indonesia               | 75,220    | 72,146      | 466,289     | 71,052         |
| 60  | Viet Nam                | 42,330    | 40,470      | 418,199     | 62,623         |
| 68  | Philippines             | 27,811    | 25,456      | 392,620     | 39,560         |
| 198 | Timor Leste             | 200       | 168         | 1,218       | 109            |
|     | **Sub-total**           | **816,299** | **772,815** | **9,421,412** | **14,25,754**  |
| 229 | Walls and Futuna        | 16        | 14          | 82          | 5              |
| 230 | Cocos (Keeling) Islands | 14        | 14          | 207         | 0              |
| 231 | South Georgia and       | 12        | 9           | 58          | 0              |
|     | the South Sandwich      |           |             |             |                |
|     | Island                  |           |             |             |                |
| 232 | Western Sahara          | 11        | 9           | 36          | 0              |
| 233 | Christmas Island        | 8         | 8           | 58          | 0              |
| 234 | Tokelau                 | 6         | 5           | 48          | 0              |
| 235 | French Southern Territories | 6   | 6          | 127         | 0              |
| 236 | Saint Pierre and        | 5         | 4           | 8           | 0              |
|     | Miquelon                |           |             |             |                |
| 237 | Bouvet Island           | 3         | 2           | 31          | 0              |
| 238 | Pitcairn               | 3         | 1           | 11          | 0              |
| 239 | Heard Island and        | 2         | 2           | 8           | 0              |
|     | McDonald Islands        |           |             |             |                |
|     | **Sub-total**           | **86**    | **74**      | **674**     | **5**          |
Citation describes the development of valid knowledge [13]. Citation gives a foothold in a scientific field to grow, which scientific growth meets the demands of filling the scientific puzzle components that are still empty [14]. In arranging the puzzle space, citations become a benchmark, and every expert/professional has done what is called the state of the art, so that the knowledge branches formed into branches: leafy to fruitful (with the presence of candidate prospective new scientists), and this function of the knowledge dissemination is carried out through scientific publications [15]. A field of science is not only soaring or excavated as a through has is difficult to reach by others, but must maintain the interrelationship between fields of science. Citation provides and requires both as an illustration of the knowledge development that applies well [16].

Citation illustrates that a scientific publication is influenced or influences other scientific publications. In other words, a scientific field is interrelated with other fields of science, or in general the birth of new scientists is caused by other scientists. This is recognized as sharing knowledge [17]. Many fields of science grow well not only because of their interests in society, but also because of citation that provokes new issues to be discussed on various occasions [18]. Therefore, the knowledge management implies that a good and bad scientific publication is based on citation, especially so that there is no overlapping of knowledge development or more extreme is the occurrence of plagiarism [19]. However, citation of scientific publications is carried out by writers in two different directions. The first is citation carried out by the author on scientific publications from other authors. The second is citation carried out by the author on his own scientific publication. Of course, two directions of citation carried out by the author have different meanings in the knowledge development.

3. An approach based on Definition
Every scientific publication (a paper in proceeding, an articles in journal, or a book, etc.) consists of several interrelated parts arranged systematically. Part by section has their respective tasks that support each other so that scientific publications become part of the development of knowledge. Certain parts that might be underestimated are references. This section of this paper aims to consider the possible values of citations carried out. Values are expressed based on generally valid consideration of the data extracted. These considerations are juxtaposed with several cases taken based on the topic of the study.

To present meaning related to citation, first stated several terms that are closely related to citations based on SCIMAGO: Document, written symbolically with $|D|$, is briefly stated as the number of documents issued during the selected year. Citation (written $s$) is the number of citations by documents issued during a particular year. Some of terms as follows, $c =$ country, $t =$ time, $cd =$ citable documents in SCIMAGO, $p =$ probability or $\%$, $ss =$ self-citation. Next, we consider $|D|, s$, and $ss$ for each country (in SCIMAGO), several journals (in SCIMAGO), and several authors (in SCOPUS).

4. Data and Assumptions
Every current scientific paper must be published online by the publisher, and will become reputable if indexed by a reputable indexing database such as Scopus [20]. Each indexer shows research performance through citation analysis. Some of the results of citation analysis, such as the number of citations in the last two years were welcomed by a database system to rank journals or proceedings [21]. Journal rating like SCIMAGO\(^1\) has the duty to provide journal/proceedings information that has quality in ranking. Data provided by SCIMAGO for example, at the time of this data collection there were $|D| =$ 48,970, 748 documents for $c =$ 239 countries registered in SCIMAGO, each of which has a number of documents referred to as country scientific outcomes.

\(^1\) https://www.scimagojr.com/
$cd_p = 92.65\%$ of the total documents or $cd = 45,373,117$ are citable documents, namely the number of documents in the form of articles, reviews, and conference papers that are exclusively considered and selected, and there are $s_t = 850, 474, 395$ citations to citable documents. Citation is expressed as a quote in years $x, x + 1, x + 2, x + 3, \ldots$ for documents issued during year $x$, and the average occurs $ss_p = 30\%$ of the number of citations or as many as $ss_t = 257, 173, 667$ are self-citations.\(^2\) In this case, self-citations (or $ss$) are intended to be the number of citations in the country received by documents issued in that country during the source year.

Table 1 shows that the 12 highest countries produce scientific publications, 7 surrounding Asian countries and Indonesia which produce scientific publications, and the 11 lowest countries produce scientific publications. Each has $cd_{st}$ as follows 31, 999, 507, 772, 815, 74, with $s_{st}$ as follows 639, 850, 637, 9, 421, 412, 674 where probability $ss_{pst}$ greater than average of $ss_{pst}$ for the highest country produce scientific publications, probability $ss_{pst}$ smaller than average of $ss_{pst}$ for 7 Asian countries and for the lowest country scientific publications. The two countries with the highest self-citations are China and United States, respectively 56\% and 46\%. This means that in the highest countries scientific publications are caused by many journals and the number of ongoing scientific meetings so that self-citations for the country itself are also high. In addition, the high level of self-citations for the country shows the publishing facilities and infrastructure in the country such as the number of available journals, or the high focus on the knowledge development in certain fields which are not carried out by other countries so that scientific publishing is only controlled by the country itself. For example, there are 13,948 journals/proceedings in the United States, 672 journals/proceedings in China, while 37 journals are only in Indonesia.

As a journal ranking system, SCIMAGO also classifies the journals/proceedings into quartiles (Q1, Q2, Q3, Q4, or non-Q). In connection with journals/proceedings independently, self-citation carried out on documents published in the journal itself or it called as self-citation journal. In general, Q1 has fewer self-citation compared to Q2, Q2 has fewer self-citation compared to Q3, and so on, but for certain journals that specifically have the specialty of accommodating scientific fields also showing self-citation for journal/proceedings also happened. Several institutions in Indonesia have produced 37 journals registered with SCIMAGO with $|D| = 4,567$ documents, consisting of 6 journals are in Q2, 14 journals are in Q3, 12 journals are in Q4, and 5 journals that do not yet have quartiles (non-Q). Each group of journals according to quartile has $|D_{stq1}| = 1,430$ documents, $|d_{stq2}| = 2,389$ documents, $|D_{stq3}| = 5,970$ documents, and $|D_{stq4}| = 151$ document. Whereas sub-totals of citable documents $cd$ are 1,442 documents, 1,674 documents, 309 with the journal self-citations are 584, 502, and 90 or the Q2 group is higher than other groups that is about 46.20\%. Self-citation journal was highest 89.84\% for one journal in group Q3, and 75.25\% for one journal in group Q2. Journal self-citations occur in journals with the highest number of documents published in the same journal, but seeing self-citation country is 15\% or less than the average 30\%, it means that citations are carried out to support scientific development in the same field as a general meaning assumption. For example, at the time of writing, in *Jurnal Pendidikan IPA Indonesia* (in Q3) there were 91 documents 128 citations and 115 self-citation journal or 89.84\%, in *Biodiversity* (in Q4) there were 201 documents 133 citations and 67 self-citation journal or 50.38\%, in *Journal of the Indonesian Tropical Animal Agriculture* (in Q4) there were 35 documents 8 citations and 4 self-citation journal or 50.00\%. Meanwhile, in the case of *International Journal on Advanced Science, Engineering and Information Technology* there were 315 documents 408 citations and 307 self-citation journal or 75.25\% which means that there is a possibility of different meaning assumptions for different journals.

The interpretation of citation in general or self-citation in particular assumes a general

\(^2\) http://www.scimagojr.com/help.php
Table 2. Self-citation of 30 authors

| doc | h_ind | s | ss | doc | h_ind | s | ss |
|-----|-------|---|----|-----|-------|---|----|
| 233 | 11    | 194 | 85 | 266 | 57    | 8989| 195 |
| 202 | 12    | 331 | 75 | 156 | 14    | 738 | 94 |
| 180 | 32    | 2606| 234| 136 | 15    | 469 | 150 |
| 172 | 12    | 303 | 48 | 124 | 7     | 105 | 38 |
| 164 | 20    | 947 | 53 | 120 | 20    | 1426| 397 |
| 149 | 17    | 478 | 151| 105 | 8     | 113 | 10 |
| 143 | 11    | 228 | 119| 102 | 5     | 36  | 12 |
| 141 | 20    | 464 | 75 | 94  | 9     | 194 | 36 |
| 140 | 13    | 384 | 42 | 93  | 7     | 197 | 7  |
| 132 | 7     | 105 | 38 | 92  | 24    | 1285| 44 |

meaning related to self-citation country and self-citation journal to self-citation authors, namely citation conducted by authors on published documents in their respective names. Self-citation author has different interpretations depending on the author’s treatment of self-citation interests of author. The self-citation authors are carried out in group in their own research groups to connect one research in the same group with other research, of course each has a different objective and research output, and this aims to enable research in the same group to support each other. In research groups, authors do their own citation between groups or on their own publications to complete scientific puzzles in their research, or generally stated to fulfill their own research (knowledge development) roadmap or the self-author. However, if the self-citation author is done to increase the h-index solely by ignoring the knowledge development roadmap. The treatment of the author with the final consideration is seen as less ethical in scientific publications, although this cannot be assessed directly as such, this is caused by the content of an article with its reference not exactly the same (except direct quotations). For example, we have surveyed 30 authors from three universities. These authors have written the highest number of documents in Scopus on behalf of each university. From each author, the citation of each document is in accordance with the author’s h-index, namely the number of citations ≥ h-index. The number of self-citations performed by the authors varies considerably between 0% to 100%. The number of documents (doc), h-index (h_ind), and citations (s), and self-citations (ss) of each author like in Table 2, and there are authors who make the highest self-citations up to 52.19% or multiples of self-citations as much as ss/doc = 333.61%, i.e. overall documents of author have cited more than 3 self-citations per documents. Overall for all documents, there are 10.40% documents of the authors in average doing self-citations.

5. Conclusion

Citation has been stated in three categories, such as self-citation, yaitu self-citation country, self-citation journal, and self-citation author. There are several semantic assumptions derived from self-citation interests that apply based on the author. The high level of self-citation country is based on the development of good knowledge for a country both in terms of public facilities and infrastructures as well as special studies, and this assumption has an influence on the need for self-citation journal and self-citation author described through h-index. Some of these assumptions requires further verification, which is a further study in other work papers.

References

[1] Syah R, Nuradi T E, Nasution M K M 2017 A framework to apply ICT for bequeathing the cultural heritage to next generation Journal of Physics: Conference Series 801(1).

[2] Bornmann L, Daniel H-D 2005 Does the h-index for ranking of scientists really work? Scientometrics 65(3).
[3] Loannidis J P A 2015 A generalized view of self-citation: Direct, co-author, collaborative, and coercive inducef self-citation Journal of Psychosomatic Research 78(1).
[4] Nasution M K M 2018 Indonesia knowledge dissemination: A snapshot Journal of Physics: Conference Series 978(1).
[5] Nasution M K M 2018 No research without publication: Early mining Journal of Physics: Conference Series 978(1).
[6] Thompson G, Yiyun Y 1991 Evaluation in the reporting verbs used in academic papers Applied Linguistics 12(4).
[7] Nasution M K M 2016 Social network mining (SNM): a definition of relation between the resources and SNA International Journal on Advanced Science Engineering Information Technology 6(6).
[8] Nasution M K M 2017 Cara menulis karya ilmiah Teknik Penulisan Karya Ilmiah, Bagian 2.
[9] Nasution M K M 2017 Merancang pernyataan masalah Desain Riset, Bagian 2.
[10] Nasution M K M 2017 Penelaahan literatur Teknik Penulisan Karya Ilmiah, Bagian 3.
[11] Popescu M, Hristea F 2011 State of the art versus classical clustering for unsupervised word sense disambiguation Artif. Intell. Rev. 35.
[12] Abt H A 2000 Do important papers produce high citation counts? Scientometrics 48(1).
[13] Lindsey D 1989 Using citation counts as a measure of quality in science measuring what’s measurable rather than what’s valid Scientometrics 15(3-4).
[14] Larsen P, Ins M V 2010 The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index Scientometrics 84(3).
[15] Mommers L, Voermans W, Koelewijn W, Kielman H 2009 Understanding the law: improving legal knowledge dissemination by translating the contents of formal sources of law Artif. Intell. Law. 17.
[16] Upham S P, Small H 2010 Emerging research fronts in science and technology: patterns of new knowledge development Scientometrics 83.
[17] Nasution M K M to print Theory of sharing knowledge: an introduction Bulletin of Mathematics.
[18] Nasution M K M 2017 Metodologi riset Desain Riset, Bagian 4.
[19] Martin B R 2013 Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment Research Policy 42(5).
[20] Moya-Anegón F D, Chinchilla-Rodríguez Z, Vargas-Quesada B, Corera-Álvarez, Muñoz-Fernández F J, González-Molina A, Herrero-Solana V 2007 Coverage analysis of Scopus: a journal metric approach Scientometrics 73(1).
[21] Bar-Han J 2010 Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar Scientometrics 82.