

Article

Relationship of Google Scholar Versions and Paper Citations

Marjan Mohammadjafari ^{1*}, Hadi Salehi ², Kaveh Bakhtiyari ^{3,4}, Nader Ale Ebrahim ⁵, Mahmoud Danaee ⁶, Masoud Shakiba ⁷, Masoomeh Shahbazi-Moghadam ⁸ and Azam Zavvari ⁹

¹ Department of Industrial Engineering, Kerman Branch, Islamic Azad University, Kerman, Iran

² Faculty of Literature and Humanities, Najafabad Branch, Islamic Azad University, Najafabad, Isfahan, Iran

³ Department of Computer & Cognitive Science, Faculty of Engineering, University of Duisburg-Essen, Duisburg 47057, Germany

⁴ Department of Electrical, Electronics and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia

⁵ Centre of Research Services, Institute of Research Management and Monitoring (IPPP), University of Malaya, 50603 Kuala Lumpur, Malaysia; aleebrahim@um.edu.my

⁶ Academic Enhancement and Leadership Development Centre (ADeC), University of Malaya, 50603 Kuala Lumpur, Malaysia

⁷ Manukau Institute of Technology, Auckland 2023, New Zealand

⁸ Perdana School of Science, Technology and Innovation Policy, Universiti Teknologi, 81310 Skudai, Johor, Malaysia

⁹ Center for Software Technology and Management, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

* Correspondence: marjan_mohamadjafari@yahoo.com; Tel.: +60-379-676-289

Abstract: The number of citations that a paper has received is the most commonly used indicator to measure the quality of research. Researchers, journals, and universities want to receive more citations for their scholarly publications to increase their h-index, impact factor, and ranking respectively. In this paper, we tried to analyse the effect of the number of available Google Scholar versions of a paper on citations count. We analyzed 10,162 papers which are published in Scopus database in year 2010 by Malaysian top five universities. Then we developed a software to collect the number of citations and versions of each paper from Google Scholar automatically. The result of Spearman correlation coefficient revealed that there is positive significant association between the number of Google Scholar versions of a paper and the number of times a paper has been cited.

Keywords: h-index; citations; published version; Scopus database; highly cited paper; bibliometrics

1. Introduction

Jorge E. Hirsch in [1] proposed the Hirsch-index which is commonly abbreviated as the H-index. The h-index is an index that attempts to measure the productivity and cumulative impact of a researcher's published work by looking at the distribution of citations the work has received [2, 3]. Another common indicator which measuring the reputation and academic standard of a journal is so-called 'Impact Factor', which, with some qualifications, is the average number of citations for papers published in a particular journal [4]. Impact Factor is obtained as the ratio of the total number of citations received by the papers published in the journal to the number of papers published in the journal [5, 6]. A majority of world university ranking are also rely on paper citations. So, receiving more citations is very important for authors, journals, and university to get high h-index, impact factor, and world ranking [7, 8]. In this research, we analyse the effect of the number of available Google Scholar versions of a paper from the web on receiving citations of the paper. We targeted to analyze all of published papers in year 2010 related to five top university of Malaysia which appear

in the Scopus database. To achieve on this purpose, 10,162 papers which are published in Scopus database in year 2010 are selected. Then we developed software to collect the number of citations and versions of each paper from Google Scholar automatically.

2. Definition of Citation

A bibliographic citation is a reference to a book, article, web page, or other published item. Citations should supply detail to identify the item uniquely [9, 10]. Citation is a reference to a published or unpublished source. Citing sources points the way for other scholars [11].

Important Purposes Citation:

- To avoiding plagiarism or support of academic honesty [12].
- To attribute prior or unoriginal work and ideas to the correct sources [13].
- To allow the reader to determine independently whether the referenced material supports the author's argument in the claimed way
- To help the reader gauge the strength and validity of the material the author has used.

3. Number of Versions for a Paper

Publishing a research paper in a scholarly journal is necessary but not sufficient for receiving citations in the future. We need to ensure that the paper is visible to the relevant users and authors. When the authors published a paper, the publisher will put the published version of the paper to own website and repository. This means we have a product, and we also have one shop then if somebody wants to have our product must come into our shop and buy it. But if we have so many versions then we can find more customers. For example, somebody made a pen and put on the one shop to sell it, beside somebody else made another pen and put on the 20 shops to sell it. It is appeared that the pen in the 20 shops is more visible for customers and then this pen will sell more.

The question is that, due to copyright roles how we can publish our paper in more than one journal to get more websites for advertise our paper. Actually, no need to publish in more than one journal but the authors can use some tools that help in enhancing the visibility and readership of research papers. Effective use of these tools can result in increased citations and, thus, improve the h-index of the author and journal impact factor. Here is a sample of tools to increase the visibility of one's published papers.

4. Strategies for Enhancing the Impact of Research Dissemination

- Submit the manuscript to a digital subject repository.
- Submit the manuscript to an institutional repository.
- Set up a web site devoted to the research project and post manuscripts of publications and conference abstracts [14].
- Take advantage of SEO (search engine optimization).
- Present preliminary research findings at a meeting or conference.
- Follow up preliminary research findings presented at a meeting or conference with a published manuscript [15].
- Consider submitting the same article to a journal in a different language as a "secondary publication."
- Start a blog devoted to the research project [16].
- Contribute to Wikipedia.
- Contribute to a social network [17].

5. Methodology

In this research, five research universities of Malaysia namely University of Malaya (UM) , Universiti Putra Malaysia (UPM), Universiti Kebangsaan Malaysia (UKM), Universiti Sains Malaysia (USM) and Universiti Teknologi Malaysia (UTM) were selected to conduct this research. We collected 10,162 papers related to year 2010 from the Scopus database, and the extraction process for collecting these papers was done in 13 July 2013 11:00 AM (UTC +8:00) for 2 hours. The process of data collection is shown in the Figures 1-7.

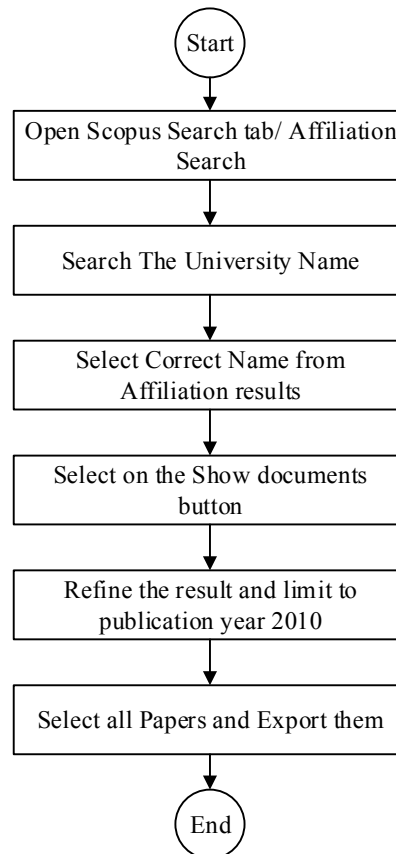


Figure1: Process of data collection from Scopus database

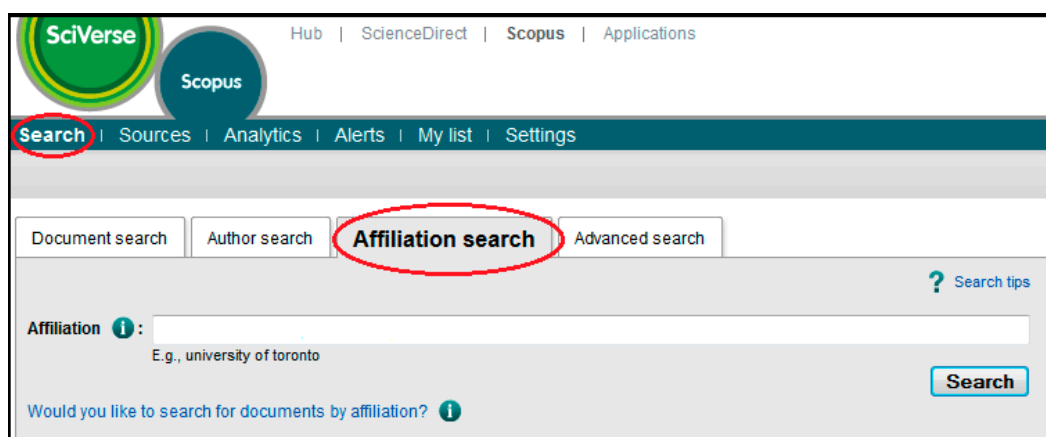
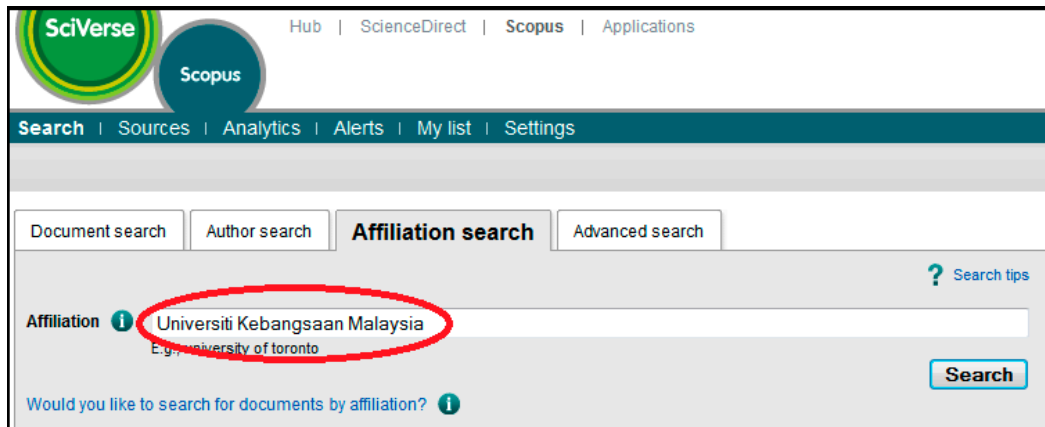


Figure 2: The capture of affiliation search in Scopus database



SciVerse | Hub | ScienceDirect | Scopus | Applications

Search | Sources | Analytics | Alerts | My list | Settings

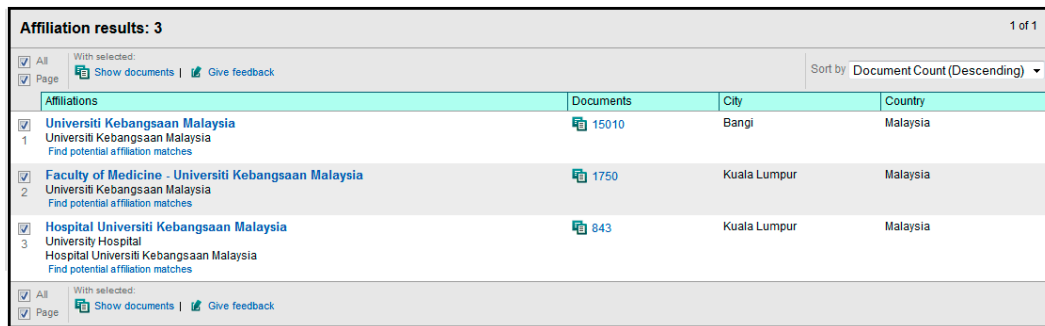
Document search | Author search | **Affiliation search** | Advanced search

Affiliation *i* [? Search tips](#)

E.g., university of toronto

Would you like to search for documents by affiliation? *i*

Figure 3: The capture of affiliation search in Scopus database



Affiliation results: 3 1 of 1

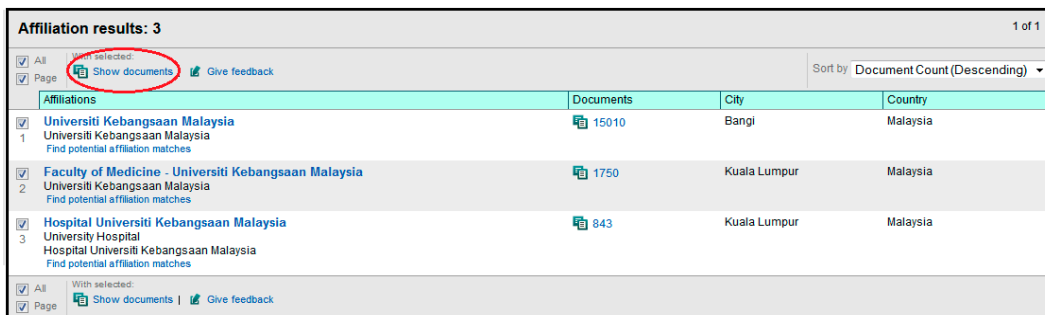
All With selected:

Page Show documents | Sort by: Document Count (Descending) ▾

Affiliations	Documents	City	Country
<input checked="" type="checkbox"/> 1 Universiti Kebangsaan Malaysia Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="15010"/>	Bangi	Malaysia
<input checked="" type="checkbox"/> 2 Faculty of Medicine - Universiti Kebangsaan Malaysia Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="1750"/>	Kuala Lumpur	Malaysia
<input checked="" type="checkbox"/> 3 Hospital Universiti Kebangsaan Malaysia University Hospital Hospital Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="843"/>	Kuala Lumpur	Malaysia

All With selected:

Figure 4: The capture of affiliation search result in Scopus database



Affiliation results: 3 1 of 1

All With selected:

Page Show documents | Sort by: Document Count (Descending) ▾

Affiliations	Documents	City	Country
<input checked="" type="checkbox"/> 1 Universiti Kebangsaan Malaysia Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="15010"/>	Bangi	Malaysia
<input checked="" type="checkbox"/> 2 Faculty of Medicine - Universiti Kebangsaan Malaysia Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="1750"/>	Kuala Lumpur	Malaysia
<input checked="" type="checkbox"/> 3 Hospital Universiti Kebangsaan Malaysia University Hospital Hospital Universiti Kebangsaan Malaysia Find potential affiliation matches	<input type="button" value="843"/>	Kuala Lumpur	Malaysia

All With selected:

Figure 5: The capture of refine affiliation to show documents in Scopus database

SciVerse Scopus Hub | ScienceDirect | Scopus | Applications

Search | Sources | Analytics | Alerts | My list | Settings Live Chat | Help | T

Quick Search Search

Your query: (AF-ID("Universiti Kebangsaan Malaysia" 60001821) OR AF-ID("Faculty of Medicine - Universiti Kebangsaan Malaysia" 60000968) OR AF-ID("Hospital Universiti Kebangsaan Malaysia" 60029395)) AND (LIMIT-TO(PUBYEAR, 2010))

Edit | Save | Set alert | Set feed

View secondary documents | Go to results: 9 Web | 217 Patent

1,960 document results | Analyze results | Show all abstracts

Search within results Search

Refine results

Limit to Exclude

Year 2010 (1,960)

Author Name

Document title	Author(s)	Date
1 Analysis of the comprehensibility of chemical hazard communication tools at the industrial workplace	Ta, G.C., Mokhtar, M.B., MOKHTAR, H.A.B.M., Ismail, A.B., ABU YAZID, M.F.B.H.	2010
2 An efficient Burst Packing algorithm for OFDMA systems	Eshanta, O.M., Ismail, M., Jumari, K.	2010

Figure 6: The capture of refine documents for year 2010 in Scopus database

SciVerse Scopus Hub | ScienceDirect | Scopus | Applications

Search | Sources | Analytics | Alerts | My list | Settings Live Chat | Help | T

Quick Search Search

Your query: (AF-ID("Universiti Kebangsaan Malaysia" 60001821) OR AF-ID("Faculty of Medicine - Universiti Kebangsaan Malaysia" 60000968) OR AF-ID("Hospital Universiti Kebangsaan Malaysia" 60029395)) AND (LIMIT-TO(PUBYEAR, 2010))

Edit | Save | Set alert | Set feed

View secondary documents | Go to results: 9 Web | 217 Patent

1,960 document results | Analyze results | Show all abstracts

Search within results Search

Refine results

Limit to Exclude

Year 2010 (1,960)

Author Name

Document title	Author(s)	Date
1 Analysis of the comprehensibility of chemical hazard communication tools at the industrial workplace	Ta, G.C., Mokhtar, M.B., MOKHTAR, H.A.B.M., Ismail, A.B., ABU YAZID, M.F.B.H.	2010
2 An efficient Burst Packing algorithm for OFDMA systems	Eshanta, O.M., Ismail, M., Jumari, K.	2010

Figure 7: The capture of export results in Scopus database

To collect the number of citations and versions of these articles, Google Scholar search engine was used. We decided to focus on this tool because of its popularity and ability to provide a simple way to find the citations of articles. Also, Google Scholar database covers more resources and it reflects more versions and citations in comparison with the other databases such as ISI Thomson Reuters or Scopus. Therefore, we developed a software to collect the number of citations and versions of each paper from Google Scholar automatically.

All records should have been processed for the number of citations and versions in only a day. Because each day, new citations and versions might be created which results the incompetency in our data analysis. In order to overcome this issue, a server-based software application was developed to retrieve citations and versions. ASP.NET platform was selected to software development, and launched on a high-speed and band-width server to be able to process all these 10,000+ records in few hours.

Software Algorithm

It searched every single title in Google Scholar in 2 times. The first time with quotation marks ("") and the second time without quotation marks. In the resulted page of Google Scholar, the titles and description might be included with some HTML tags as below:

 : For the keywords match the search query, they would be bold to show the matched title with the keywords.

<i> </i>: This tag was also found in few titles in Google Scholar search results.

 : Those titles with the superscripts and subscripts (e.g. Chemical formulas) consist of these tags to show the titles properly.

In order to extract and find the correct matching title in Google Scholar, all these HTML tags were removed from the titles. But still there was another challenge, and it was the different spacing. Some of the titles extracted from Scopus were 1 or 2 spaces different with those indexed in Google Scholar. So, after removal of all tags from titles, all spaces were also removed to find the correct match of the paper in Google Scholar results. In some cases, there were more than two items matched with the full title, and then in this case, the year and the authors' names of the published data were matched to find the relevant record.

If the title, were found, then the number of citations and versions were extracted from the page and it was updated in the database; and if it was not found, it would be marked as "not found" in the database. The whole extraction process was done in 15 July 2013 12:00 AM (UTC +8:00) for 4 hours. After completion of the data extraction, those not found records, were also checked manually to make sure that system and data analysis have minimum incomplete data and no record has been missed on Google Scholar. The structural procedure is visualized in the Figure 8.

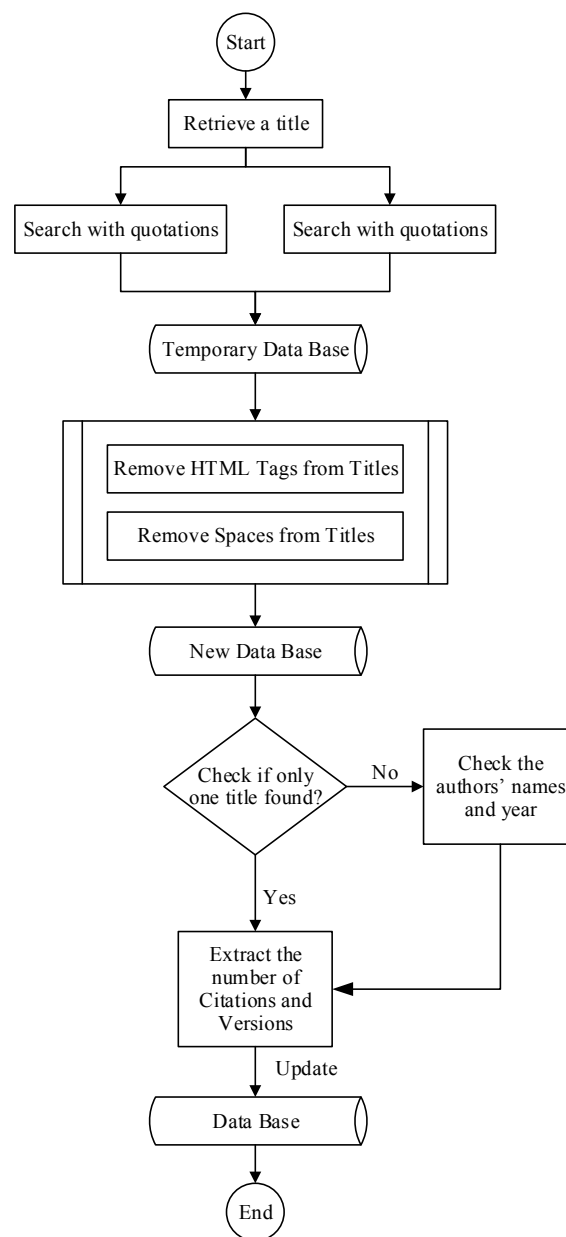


Figure 8: The structure of developed software

6. Data Analysis

As both number of citations and number of versions were not distributed normally, a non-parametric method, was applied to test the differences among top five universities. Table 1 shows the collected data of five top Malaysian research universities for year 2010 from Scopus database. The result of spearman correlation coefficient revealed that there was positive significant association between number of citation and number of versions for different universities publication. The overall correlation coefficient was a moderate and positive ($r = 0.431$, $p < 0.01$). Relationship between number of citation and number of version shows in Table 2.

Table 1: Total number of publications and citations of five top universities of Malaysia

University	Number of Publication	Number of Citations
UKM	1961	7243
UM	2276	11118
UPM	2061	9251
USM	2480	12575
UTM	1384	5081
Total	10162	45268

Table 2: Correlation between number of citations and number of versions

Number of Citation	Number of Versions					
	UKM	UM	UPM	USM	UTM	Total
	.413**	.478**	.403**	.427**	.405**	.431**

***. Correlation is significant at the 0.01 level (2-tailed).*

6.1. Comparison among 5 Malaysian top universities for number of citations and number of versions

As both number of citations and number of versions were not distributed normally, Kruskal Wallis test, a non-parametric method, was applied to test the differences among these universities. The results revealed that there were significant differences among these 5 universities for both number of citation and version (Table 3).

Table 3: The results of Kruskal Wallis Test for comparison among universities

	Number of citation	Number of versions
Chi-square	51.786	21.816
df	4	4
P value	<0.01	<0.01

Figures 9 and 10 showed the differences for the average number of citation and version among universities. The highest mean number of citation per article observed for USM ($M = 5.07$) and UM ($M = 4.88$) respectively while UTM ($M = 3.69$) and UKM ($M = 3.67$) had the lowest mean number of citation per article.

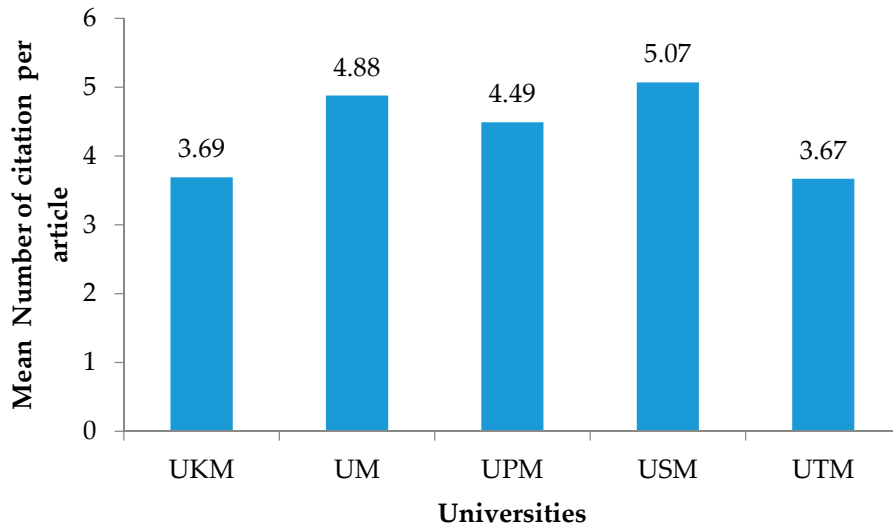


Figure 9: The differences for the average number of citations among universities

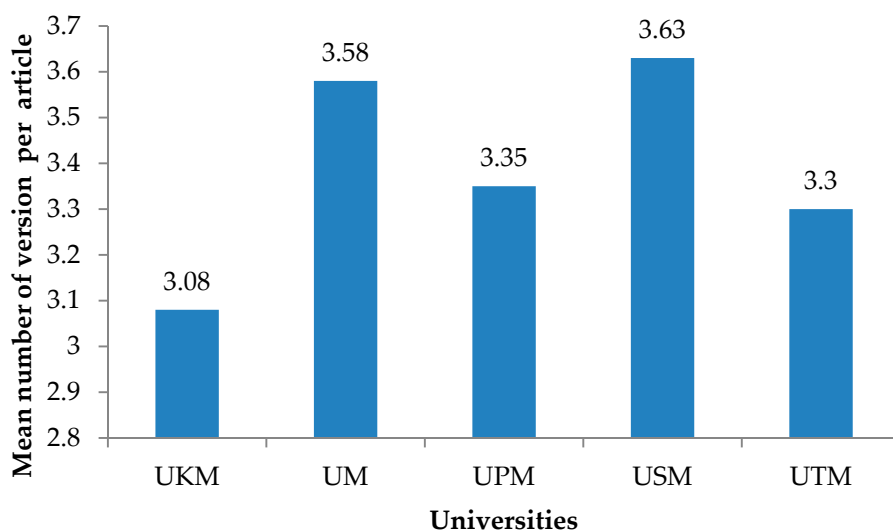


Figure 10: The differences for the average number of version among universities

6.2. Relationship between type of document and type of university for number of publication

The frequency of different type of document in each university was calculated. Table 4 showed the pattern of publication in each university based on the document type. To study the relationship between type of document and universities a chi square test was applied the result of Fisher exact test revealed that there was a significant relationship between type of publication and university (Table 5).

Table 4: Number of publication Relationship between type of document and type of university

Type of document	UKM		UM		UPM		USM		UTM	
	Count	%	Count	%	Count	%	Count	%	Count	%
Article	1260	64.30%	1767	77.60%	1673	81.20%	1915	77.20%	632	45.70%
Article in Press	5	0.30%	6	0.30%	4	0.20%	5	0.20%	1	0.10%
Conference Paper	644	32.80%	395	17.40%	306	14.80%	463	18.70%	728	52.60%
Editorial	1	0.10%	14	0.60%	7	0.30%	10	0.40%	8	0.60%
Erratum	1	0.10%	3	0.10%	2	0.10%	3	0.10%	1	0.10%
Letter	11	0.60%	21	0.90%	3	0.10%	12	0.50%	0	0.00%
Note	3	0.20%	8	0.40%	1	0.00%	3	0.10%	1	0.10%
Review	36	1.80%	61	2.70%	65	3.20%	67	2.70%	11	0.80%
Short Survey	0	0.00%	1	0.005%	0	0.00%	2	0.10%	2	0.10%

Table 5: Type of publication and university

	Value	df	P value
Pearson Chi-Square	915.163	32	<0.01
Fisher's Exact Test	853.92	--	<0.01

6.3. Comparison among different types of publication for number of citation and number of versions

Kruskal Wallis test was applied to test the differences for number of citation and number of versions among different types of publication. The results revealed that there were significant differences for both number of citation and version (Table 6).

Table 6: The results of Kruskal Wallis Test for comparison among different types of publication

	Number of citations	Number of versions
Chi-square	691.304	584.03
df	8	8
P value	<0.01	<0.01

Figure 11 shows the differences for the average number of citations and versions among universities the highest mean for citations per article observed for review article (M=16.07) and the lowest was for editorial (M=0.45).

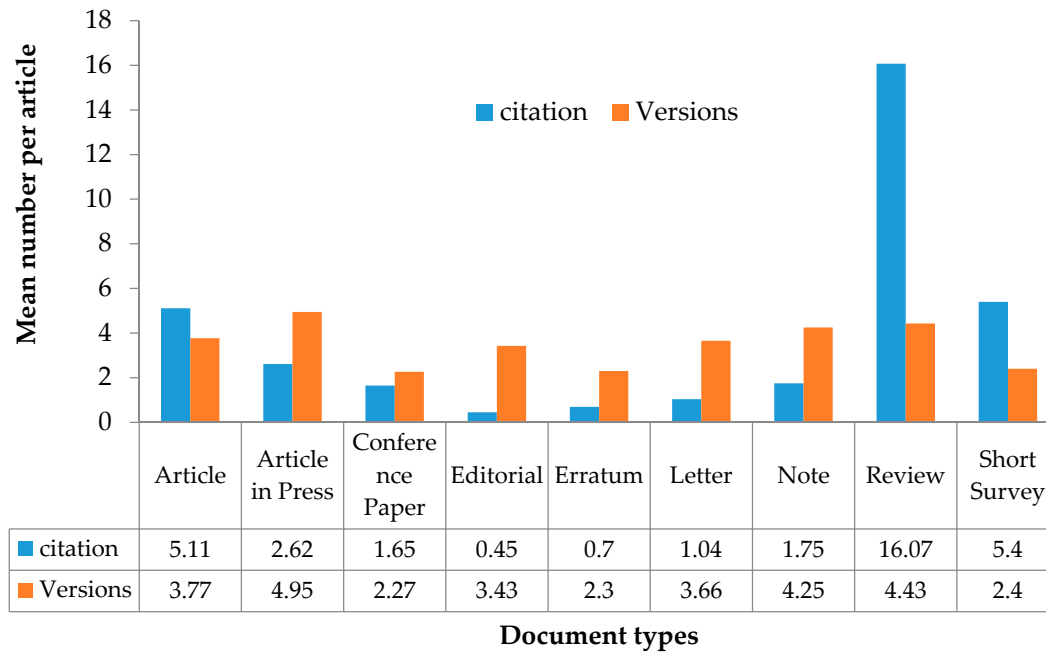


Figure 11: The differences for the average number of citations and versions among universities

7. Conclusion

In this paper, we analysed the effect of the number of available Google Scholar versions of a paper from the web on receiving citations for the paper. We targeted to analyze all of published papers in year 2010 related to five top university of Malaysia which appear in the Scopus database. To achieve on this purpose, 10,162 papers which are published in Scopus database in year 2010 are selected. Then we developed a software to collect the number of citations and versions of each paper from Google Scholar automatically. Since there is a strong association between the number of Google Scholar versions of a paper and the number of times a paper has been cited, we encourage the researchers to increase the number of paper versions by depositing in different open access repositories.

Author Contributions: All authors checked manually on Google Scholar for “not found records” which returned by the software. Nader Ale Ebrahim did the data collection and final touch up of the paper. While, Marjan Mohammadjafari; wrote the initial draft of the paper, Hadi Salehi; revised it, and Mahmoud Danaee analyzed the data. Kaveh Bakhtiyari was the software programmer. Masoud Shakiba, Azam Zavvari, and Masoomah Shahbazi-Moghadam contributed to clean the data and wrote some parts of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hirsch, J.E., *An index to quantify an individual's scientific research output*. Proceedings Of The National Academy Of Sciences Of The United States Of America, 2005. **102**(46): p. 16569-16572.
2. Farhadi, H., et al., *Does it Matter Which Citation Tool is Used to Compare the h-index of a Group of Highly Cited Researchers?* Australian Journal of Basic and Applied Sciences, 2013. **7**(4): p. 198-202.
3. Karpagam, R., et al., *Mapping of nanoscience and nanotechnology research in India: a scientometric analysis, 1990–2009*. Scientometrics, 2011. **89**(2): p. 501-522.
4. Fooladi, M., et al., *Do Criticisms Overcome the Praises of Journal Impact Factor?* Asian Social Science, 2013. **9**(5): p. 176-182.
5. Bickers, D.R. and R.L. Modlin, *A Review of the Journal of Investigative Dermatology's Most Cited Publications over the Past 25 Years and the Use of Developing Bibliometric Methodologies to Assess Journal Quality*. Journal of Investigative Dermatology, 2012. **132**(3): p. 1050-1060.
6. Day, R. and B. Gastel, *How to write and publish a scientific paper*. 2012: Cambridge University Press.
7. Ale Ebrahim, N., et al., *Visibility and Citation Impact*. International Education Studies, 2014. **7**(4): p. 120-125.
8. Ale Ebrahim, N., et al., *Effective Strategies for Increasing Citation Frequency*. International Education Studies, 2013. **6**(11): p. 93-99.
9. Goodrum, A.A., et al., *Scholarly publishing in the Internet age: a citation analysis of computer science literature*. Information Processing & Management, 2001. **37**(5): p. 661-675.
10. Meho, L.I. and K. Yang, *Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar*. Journal of the american society for information science and technology, 2007. **58**(13): p. 2105-2125.
11. MIT. *Academic Integrity at MIT, A Handbook for Students*. 2013; Available from: <https://integrity.mit.edu/citing-your-sources/avoiding-plagiarism-cite-your-source>.
12. Roig, M., *Avoiding plagiarism, self-plagiarism, and other questionable writing practices: A guide to ethical writing*. 2011.
13. Rubio, C.C., *Colombia's Poetic World of Authors' Moral Rights: Considerations on Imprisoning a Professor for Plagiarism*. 2012.
14. Berry, R., *The research project: how to write it*. 2013: Routledge.
15. Gur, T., *The Contribution of Special Education Students' Clinical Work to their Sense of Self Efficacy: A Preliminary Research*. Procedia-Social and Behavioral Sciences, 2012. **69**: p. 997-1003.
16. Wilson, J., *Essentials of business research: A guide to doing your research project*. 2014: Sage.
17. McKenzie, G., et al. *What, When and Where: The real-world activities that contribute to online social networking posts*. in *International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY)*, Leuven, Belgium, edited by Peter Mooney, and Karl Rehl. 2013.