Producing Top LIS Journal Paper List Based on the Yearly Citation Growth Rate

연간 인용 횟수 증가율에 기반한 문헌정보학 학술지 논문 목록의 순위화에 관한 연구

Eungi Kim*

<Contents>

ABSTRACT

This study proposes a novel method to rank highly-cited papers that incorporate the likelihood of receiving future citations. Instead of using the total citation count, the proposed method ranks most-papers based on the yearly citation growth rate (YCGR). The rank of YCGR can be obtained by calculating the average ranks of five individual citation related variables: 1) Total Citation Count, 2) Leftside-Slope, 3) Publication Year, 4) Peak Year, and 5) Rightside-Slope. To empirically test the proposed method, yearly citation counts with other relevant bibliographic records of the 50 most-cited papers in Library and Information Science (LIS) journals used in the study conducted by Walters and Wilder were collected from the Scopus database for the years 1996 to 2016. The result indicated that the YCGR appears to reflect the degree to which the paper is likely to receive future citations, and the ranked list based YCGR offered an alternative viewing feature of the highly-cited papers in LIS. Although more empirical analyses are needed, the rank based on YCGR in conjunction with variables related to YCGR can be used as an alternative method in recognizing influential papers in LIS.

Keywords: yearly citation count, slope, most-cited papers, citation trend, the ranking, YCGR

초 록

본 연구는 기존의 인용횟수를 기반으로 순위매김을 하던 방식과 달리 연도별 인용횟수증가율 (YCGR) 을 기반으로 하는 특정 방법으로 순위매김을 하여 추후의 인용가능성까지 아우르는 새로운 순위매김 방법론을 제안한다. YCGR 기반 순위매김은 논문의 인용횟수만을 적용하지 않고 총 인용문, 왼쪽경사, 출판연도, 피크연도 및 오른쪽경사와 같은 5 가지 개별인용 관련 변수의 평균값을 계산하여 순위를 정한다. 이 연구를 수행하기 위해 Scopus 데이터베이스를 이용하여 1996년에서 2016년까지 Walters와 Wilder가 발표한 문헌정보학 저널을 기반으로 삼아 가장 많이 인용된 논문 50편을 수집하였다. 이 논문들을 앞서 언급한 YCGR 측정요건들에 맞게 설정한 기울기를 측정하여 순위 매김을 하였다. 결과적으로 YCGR 순위는 논문이 추후 인용 받을 가능성의 정도를 반영하는 것으로 보이며 또한 문헌정보학에서 많이 인용된 논문들의 특징과 추세를 살펴볼 수 있는 대안적 기능도 제공했다. 더 많은 경험적 분석이 필요하지만 YCGR에 기반한 순위는 다양한 연구영역의 최다 인용된 논문들 가운데서 주목할 만한 논문을 찾는 데에도 유용하게 쓰여질 수 있을 것으로 보인다.

키워드: 연간 인용수, 기울기, 가장 많이 인용된 논문, 인용 추세, 순위, 선영 회귀, YCGR

* 계명대학교 문헌정보과 조교수(egkim@gw.kmu.ac.kr)
•논문접수: 2018년 2월 20일 •최초심사: 2018년 3월 1일 •게재확정: 2018년 6월 18일
•한국도서관·정보학회지 49(2), 245-266, 2018. [http://dx.doi.org/10.16981/kliss.49.201806.245]

Ⅰ. Introduction

It was long believed that the most-cited papers based on the total citation count could effectively represent a research domain. Dating back to more than 30 years, Garfield (1987) has shown that the most-cited papers can be used to represent some noteworthy papers in the medical field. Since the most-cited papers would have had a greater attention from other researchers, it is reasonable to assume that the most-cited papers are likely to have a greater influence on the topical areas with which the paper is associated than the papers receiving little or no citation.

Despite the popularity, one weakness of recognizing the papers based on the total citation is that the most-cited papers do not indicate the likelihood of receiving future citations. Hence, the ranked list may not reflect the growth rate of citation pertaining to the paper. Simply producing a list of papers based on the total citation is often trivial. The bibliographic database providers already offer such features to sort the retrieved data based on the total citation count. Rather than using the total citation count, the method described in this paper uses the trend of receiving citation based on the yearly citation.

Producing such a list of papers that incorporates the likelihood of receiving future citations can be beneficial for those researchers who attempt to examine the subject areas, discipline, and topical areas of research. Because all papers can be categorized according to subject areas in one way or another, citation count of papers can be examined according to subject areas. A subject area comprised of papers receiving a higher rate of citation may indicate a fast-growing subject. Furthermore, viewing papers that are likely to receive a higher rate of citations can aid in finding potential topic areas of research and may provide assurance that their research interests are in line with the most current interests in an academic community. Despite many works on citation patterns, previous studies have not adequately focused on providing alternative methods to identify the most influential papers in LIS.

To this end, the aim of this paper is to investigate an alternative way to rank papers that would reflect the potential likelihood of receiving more citation in the coming years. The proposed method uses the yearly citation growth rate (YCGR). YCGR incorporates citation related variables: publication year, total citation rank, the year where yearly citation peaked, and the slope of yearly citation trend. Specifically, YCGR is an average value of these variables, and, in essence, it reflects the growth rate of receiving attention from other papers. In the remaining parts, this paper describes a novel approach in measuring YCGR and the result of ranking the most-papers based on YCGR.

Ⅱ. Related Studies

By using the most-cited papers, previous works identified various prominent theories, applications, and practices in various fields of studies. The most-cited paper method has been particularly utilized to demonstrate the important research works in the medical field. Jafarzadeh et al. (2015) identified the 100 most-cited articles in the field of dental, oral, and maxillofacial traumatology over the last 64 years. Feijoo et al. (2014) demonstrated that the most-cited articles in dentistry comprised of clinical type studies that usually cover specific cases or are based on narrative reviews or expert opinions. Brinjikji, Klunder, and Kallmes (2013) analyzed the most-cited medical imaging articles from the year 2000 or later.

Pertaining to the field of LIS, there have been attempts to highlight important works based on citation counts. Blessinger and Hrycaj (2010) revealed characteristics of 32 highly-cited articles published between 1968 and 2008 using 28 journals listed in *Journal Citation Reports* (JCR). They reported that two journals that published a substantial number of the most-cited articles were *Journal of the American Society for Information Science* (38%) and *Journal of Documentation* (31%). In terms of authorship, institutions, and network structures, Bauer, Leydesdorff, and Lutz Bornmann (2016) examined most-cited LIS papers published between 2002 and 2012 by using the Web of Science (WOS) database. The authors pointed out that the most-cited top 1% papers tended to reflect the current LIS research trends. Ivanović and Ho (2010) showed that the most-cited paper in LIS was the article entitled ‘Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology’ written by Davis (1989). Since the paper written by Davis was published almost three decades ago, extracting the most-cited papers without considering the publication date may be problematic in finding a list of most influential papers for students and researchers of LIS.

Numerous works have been conducted without focusing on a specific field of research. Rather, some researchers focused on finding distinctive characteristics of the most-cited papers. For instance, Aksnes (2005) pointed out that the most-cited papers tend to receive more citations from other papers due to increased visibility through citations. Similarly, using papers published in Science, Newman (2014) suggested that one can easily identify papers that will likely to receive more citations simply by examining on the past received citations.

Other authors focused on identifying distinctive patterns based on the yearly citation counts. These include the life cycle, peak citation year, and half-life. The life cycle of an article indicates the length of its citability since the paper was published. Walters (2011) reported that the most-cited papers tend to be cited for more than 25 years after initial publication. Half-life is the median age for which it has received citation. Based on the journals published in the Journal Citation Report, Davis (2015) found that the half-life for the scholarly literature is 6.5 years. Aversa (1985) pointed out that the most-cited papers tend to age more slowly as compared to the papers receiving few citations. Similarly, Levitt and Thelwall (2008) found that the annual citation of most-cited papers usually peaks much later than weakly cited papers. These works were based on examining the yearly citation. These findings were useful in understanding the citation pattern of most-cited papers.

Some related studies also focused on predicting future citation counts based on the existing citation related data. Yan et al.'s (2011) approach included a number of machine learning related features. The author's approach included an extensive number of variables: the topic, diversity, recency, h-index, author rank, productivity, sociality, authority, venue rank, and venue centrality. They constructed an elaborate algorithm to predict the future citation count. Lokker et al. (2008) used 27 variables including the authors and the number of pages of the paper. In predicting future citation, the authors used an extensive number of features (variables), including the required machine learning algorithms. While these types of works may work quite accurately, in these types of machine learning approaches, the process of determining the likelihood of receiving more citation in the future is not intuitive for the user.

To find influential papers, the likelihood of receiving future citations could be incorporated. This requires several factors such as publication year, the year where yearly citation peaked, and slopes of the trending line. With consideration of variables related to yearly citation count, this paper presents a relatively straightforward method in ranking the papers that are likely to receive more citations in the future. Although the success rate of predicting the future citation may not be better than the machine learning based approaches, the proposed method in this paper is relatively simple and intuitive since the user can intuitively understand the effects of using a limited number of key variables related to yearly citation count. Consequently, the user would have better control in observing the citation trends of individual highly-cited papers.

Ⅲ. Methodology

This section describes the methodological procedures used in conducting this research. An appropriate dataset was needed for a variety of reasons: a) to describe the rank of the papers based on YCGR, b) to justify the proposed method using the empirical data, c) to test the applicability to the field of LIS, and d) to compare the results of the proposed method to the conventional method, which solely relies on the total citation count.

Initially, LIS journals and the journal categories were selected for creating a dataset. Walters and Wilder (2015) proposed six journal categories for LIS journals. Using their journal category, all 29 journals listed in their LIS journal category were searched with the publication period from 1996 to 2016 using *Scopus.* Although there are other LIS journals that published highly-cited papers, 29 journals appear to be a reasonable size in representing the highly-cited papers in LIS. The details of the journals and the journal categories used in this study are shown in Table 1. These journal categories were also used in the study conducted by Lee and Bak (2016). For convenience, the journals have been assigned with acronyms that reflect the subject category and the index number. In this table, the total number of documents published for each journal is also shown.

Finally, the dataset was created by extracting 50 papers using the *Scopus* database (http://www.scopus.com). Because *Scopus* provides an ability to search the records based on the citation count and other criteria, extracting only the 50 most-cited papers was necessary to conduct this study. Although the quantity of 50 papers was arbitrarily chosen, it appeared to be a reasonable number since all the papers published in these journals received a high number of citations. In the subsequent section, the downloaded 50 most-cited papers were used in describing the process of producing a ranked list based on YCGR.

<Tab. 1> Journals Used to Create the Dataset

Journal Categories	Journal Names	Journal Index	# of Docs
LIS Core Journals (CORE)	Aslib Journal of Information Management	CORE1	99
	Information Research	CORE2	823
	Journal of the American Society for Information Science – JASIST (formerly JASIS)	CORE3	3423
	Journal of Documentation	CORE4	998
	Journal of Librarianship and Information Science	CORE5	466
	Knowledge Organization	CORE6	515
	Library and Information Science Research	CORE7	635
	Library Quarterly	CORE8	419
	Libri	CORE9	552
Practice-Oriented Journals (PR)	College & Research Libraries	PR1	768
	Health Information and Libraries Journal	PR2	731
	Information Technology and Libraries	PR3	529
	Journal of Academic Librarianship	PR4	1406
	Journal of the Medical Library Association*	PR5	16
	Library Collections, Acquisition and Technical Services	PR6	386
	Library Hi Tech	PR7	1024
	Library Resources & Technical Services	PR8	454
	Portal: Libraries and the Academy	PR9	553
	Serials Review	PR10	1672
Computer Science-Oriented Journals (CS)	Information Processing and Management	CS1	1403
Management-Oriented Journals (M)	Government Information Quarterly	M1	891
	Online Information Review	M2	1084
Informetrics Journals (INFOM)	Journal of Informetrics	INFOM1	700
	Scientometrics	INFOM2	3585
Other LIS Journals (OTHER)	Electronic Library	OTHER1	1419
	Information Society	OTHER2	639
	Journal of Scholarly Publishing	OTHER3	419
	Libraries & The Cultural Records*	OTHER4	1
	Library Trends	OTHER5	908

Note: The symbol "*" indicates the journals that are no longer indexed by the Scopus database.

Ⅳ. Results

1. The 50 Most-Cited Papers as Dataset

Table 2 shows the result of extracting the 50 most-cited papers. In this table, the rank of papers was determined based on the total citation count, which starts from 1263 to 80. Some notable papers that received an exceptional total citation count can be observed in terms of citation distribution. The total citation count would be plotted as an exponential line (Barabási, Song and Wang 2012; Patience et al. 2017). In Table 2, the leftmost column indicates the paper index. The paper index comprises of the rank of total citation counts along with the journal index. In the subsequent section of this paper, the most-cited papers shown in this table will be referred to this indexing scheme.

<Tab. 2> The 50 Most-cited Papers in the Dataset

Paper Index (Citation Rank-Journal Index)	Pub. Year	Title	Authors	Total Cit-ation Count
1-CORE3	2007	The link-prediction problem for social networks	Liben-Nowell D., Kleinberg J.	1263
2-M1	2001	Developing fully functional E-government: A four stage model	Layne K., Lee J.	841
3-CORE3	2009	Twitter power: Tweets as electronic word of mouth	Jansen B.J., Zhang M., Sobel K., Chowdury A.	826
4-INFOM2	2004	Citation review of Lagergren kinetic rate equation on adsorption reactions	Ho Y.-S.	804
5-CS1	2000	Real life, real users, and real needs: A study and analysis of user queries on the Web	Jansen B.J., Spink A., Saracevic T.	797
6-INFOM2	2006	Theory and practise of the g-index	Egghe L.	662
7-CORE4	1999	Models in information behaviour research	Wilson T.D.	633
8-CS1	2009	A systematic analysis of performance measures for classification tasks	Sokolova M., Lapalme G.	607
9-CORE3	2001	Searching the Web: The Public and Their Queries	Spink A., Wolfram D., Jansen M.B.J., Saracevic T.	556
10-M1	2010	Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies	Bertot J.C., Jaeger P.T., Grimes J.M.	542
11-PR2	2009	A typology of reviews: An analysis of 14 review types and associated methodologies	Grant M.J., Booth A.	528
12-CORE3	2006	CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature	Chen C.	481
13-CORE3	2010	Sentiment in short strength detection informal text	Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas A.	479
14-PR2	2007	The emerging Web 2.0 social software: An enabling suite of sociable technologies in health and health care education	Boulos M.N.K., Wheeler S.	465
15-OTHER2	2002	Information systems and developing countries: Failure, success, and local improvisations	Heeks R.	457
16-CORE3	2009	A survey of modern authorship attribution methods	Stamatatos E.	451

17-OTHER2	2004	The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines	Eppler M.J., Mengis J.	436
18-CORE3	2007	Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar	Meho L.I., Yang K.	427
19-M1	2007	Analyzing e-government research: Perspectives, philosophies, theories, methods, and practice	Heeks R., Bailur S.	414
20-CS1	2004	Centroid-based summarization of multiple documents	Radev D.R., Jing H., Styś M., Tam D.	403
21-M1	2007	E-government research: Reviewing the literature, limitations, and ways forward	Yildiz M.	402
22-CS1	2006	How are we searching the World Wide Web? A comparison of nine search engine transaction logs	Jansen B.J., Spink A.	395
23-M1	2005	E-government success factors: Mapping practical tools to theoretical foundations	Gil-García J.R., Pardo T.A.	394
24-CORE4	2008	What do citation counts measure? A review of studies on citing behavior	Bornmann L., Daniel H.	365
25-OTHER2	2003	The Digital Divide as a Complex and Dynamic Phenomenon	Van Dijk J., Hacker K.	363
26-INFOM2	2010	Software survey: VOSviewer, a computer program for bibliometric mapping	van Eck N.J., Waltman L.	359
27-CORE4	2004	Understanding inverse document frequency: On theoretical arguments for IDF	Robertson S.	355
28-INFOM2	2005	Mapping the backbone of science	Boyack K.W., Klavans R., Börner K.	353
29-CS1	2000	Probabilistic model of information retrieval: Development and comparative experiments. Part 1	Sparck Jones K., Walker S., Robertson S.E.	350
30-CORE4	1996	Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory	Ingwersen P.	348
31-CS1	2005	Co-authorship networks in the digital library research community	Liu X., Bollen J., Nelson M.L., Van De Sompel H.	342
32-CORE7	1996	Social network analysis: An approach and technique for the study of information exchange	Haythornthwaite C.	342
33-PR2	2007	Second Life: An overview of the potential of 3-D virtual worlds in medical and health education	Boulos M.N.K., Hetherington L., Wheeler S.	332
34-CORE3	2002	Judgment of information quality and cognitive authority in the Web	Rieh S.Y.	328
35-CORE3	2002	Believe it or not: Factors influencing credibility on the Web	Wathen C.N., Burkell J.	321
36-CS1	2007	Text mining techniques for patent analysis	Tseng Y.-H., Lin C.-J., Lin Y.-I.	316
37-CS1	1997	Trends in … A critical review: Information behaviour: An interdisciplinary perspective	Wilson T.D.	312
38-INFOM2	2006	Comparison of the hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups	Van Raan A.F.J.	305
39-CORE2	2002	The nonsense of 'knowledge management'	Wilson T.D.	304
40-PR5	2006	How to identify randomized controlled trials in MEDLINE: Ten years on	Glanville J.M., Lefebvre C., Miles J.N.V., Camosso-Stefinovic J.	303

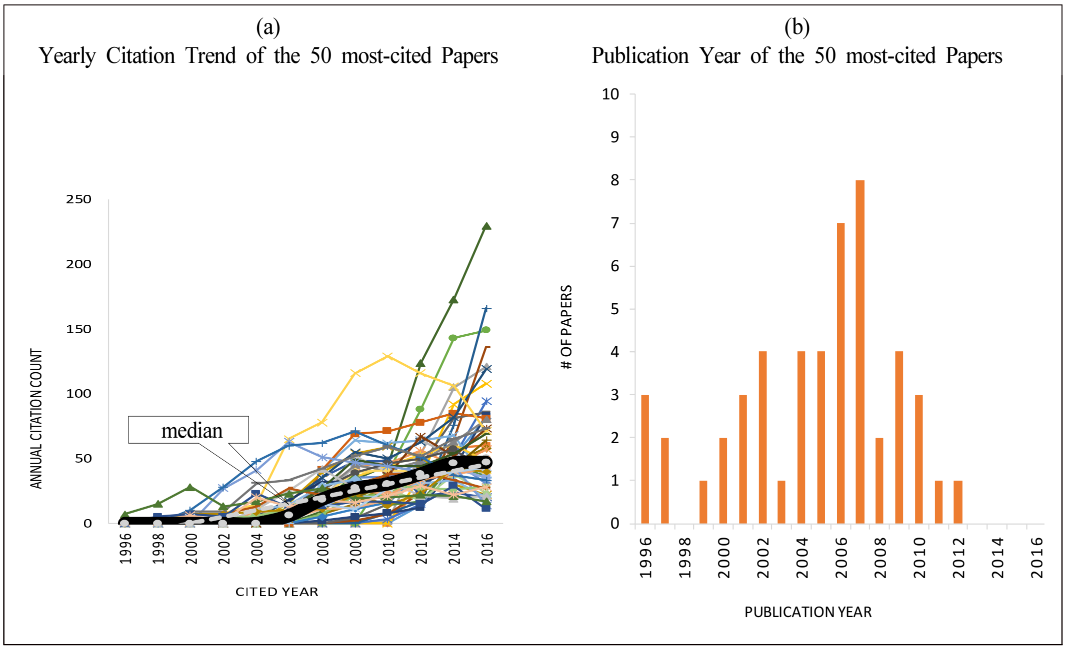
41-CORE8	1996	Modeling the information seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers	Leckie G.J., Pettigrew K.E., Sylvain C.	301
42-INFOM2	2005	Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods	Van Raan A.F.J.	298
43-CORE3	2011	Sentiment in Twitter events	Thelwall M., Buckley K., Paltoglou G.	296
44-CORE3	2008	What is user engagement? A conceptual framework for defining user engagement with technology	O'Brien H.L., Toms E.G.	290
45-INFOM2	2001	National characteristics in international scientific co-authorship relations	Glänzel W.	287
46-CORE3	2012	Sentiment strength detection for the social web	Thelwall M., Buckley K., Paltoglou G.	286
47-CORE3	2007	Making sense of credibility on the web: Models for evaluating online information and recommendations for future research	Metzger M.J.	286
48-CORE4	1997	Informetric analyses on the world wide web: Methodological approaches to 'webometrics'	Almind T.C., Ingwersen P.	283
49-M1	2006	E-government maturity models: Extension of the Layne and Lee model	Andersen K.V., Henriksen H.Z.	280
50-CORE3	2006	A framework for authorship identification of online messages: Writing-style features and classification techniques	Zheng R., Li J., Chen H., Huang Z.	280

2. Visualization of Yearly Citation Trends of the 50 Most-Cited Papers

Plotting the publication year of the most-cited papers allows us to visualize the distribution of the sampled 50 most-cited papers. Figure 1 portrays the total citation growth pattern of the top 50 most-cited papers. In Figure 1(a), a varying citation trend of all 50 individual papers is depicted. As previous studies suggested, there is a greater variability among yearly citation count (Levitt and Thelwall 2008). As depicted in the median of the yearly citation count, citation count tends to rise toward the right side of the graph. This trend appears to be general among the varying datasets based on the most-cited papers (Levitt and Thelwall 2008).

The publication year trend is shown in Figure 1(b). In this figure, we can notice the number of papers compared to publication year. As shown the years where the highest number of papers was published are 2007 and 2008. We can see that papers published after 2012 do not make into the 50 most-cited paper list.

<Fig. 1> Distribution of the Yearly Citation Count and Publication Year



3. Variables Related to Yearly Citation Count

It is useful to define variables related to the yearly citation counts in formulating YCGR. These include:

- Leftside-Slope,
- RSQ-Left,
- Rightside-Slope,
- Year-Starting
- Year-Ending
- Peak Year

Leftside-Slope is the linear regression slope from a publication year to Peak Year. Peak Year is the year in which the paper received the most citations. RSQ-left is the R-square value of the linear regression line starting from the publication year to the Peak Year. Rightside-Slope is the linear regression scope from the Peak Year to Year-Ending. Year-Ending is the right most recent year in the citation window period. Citation window period specifies the years that the paper received citation. Year-Starting is the left most recent year in the citation window period (i.e., the year that the paper has received the first citation). In this study, the regression line was calculated starting from publication year rather than the Year-Starting. Peak Year would be equal to the Year-Ending of the citation window period only if yearly citation count is the highest at the Year-Ending. For instance, for the citation window period of 1996 to 2016, the yearly citation count would be the highest, if the Peak Year is determined to be 2016.

For the most-cited papers, we can expect a general increase in the yearly citation count after the publication year. This paper shows that the rate of growth in receiving yearly citation counts can be roughly estimated by relying on simple regression lines. Although more citations will come as the time passes, based on the data, two types of Peak Years can be observed:

A) Peak Year equals Year-Ending

B) Peak Year does not equal Year-Ending

In case of A, Rightside-Slope is not applicable since the values cannot be calculated. To have a valid Rightside-Slope, another year after the Peak Year must exist. In case of B, the Rightside-Slope is relevant and can be calculated. In this type of case, two slope lines – the inclining line and the declining line – can be examined in combination in order to determine YCGR. If the Peak Year falls between the citation window period rather than the Year-Ending, then the Rightside-Slope would have 0 or negative value. Consequently, the Rightside-Slope can be incorporated in determining YCGR. A calculation of Rightside-RSQ was avoided since there would be a minimal number of data points from Peak Year to Year-Ending. However, if the Peak Year is formed early, then Rightside-RSQ can be easily included in this table.

4. Journal Categorization and Measurement of Variables

Journal categories of the 50 most-cited papers and citation related variables can be empirically analyzed. Table 3 shows journal categorization and various measurements of some citation related variables that were identified earlier. This table shows that journals varied in publishing the 50 most-cited papers. For example, 14 papers (28%) are published by CORE3 (*JASIST*) and 7 papers (14%) are published by CORE4 (*Journal of Documentation*). Evidently, some journals are able to publish a greater number of highly-cited papers than other journals. The Leftside-Slope would be almost positive for papers receiving at least one citation. In the case of the 50-most papers, the Leftside-Slope ranges from 1.04 (Citation Rank #37) to 27.88 (Citation Rank #1). The leftside-RSQ could also be taken into consideration. In the case of the 50 most-cited papers, the Leftside-RSQ ranges from .48 to .99. It should be noted that 17(34%) papers had Year-Ending 2016 as the Peak Year. Because Year-Ending 2016 does not have a Rightside-Slope due to having a positive slope, some Rightside-Slope was labeled with “N/A’.

〈Tab. 3〉 Journal Categorization and Measurement of Variables

Subject Category	Publication Year	Journal Category	Total Citation Rank	Peak Year	Leftside-Slope	Leftside-RSQ	Rightside-Slope
LIS Core Journals (CORE)	2007	CORE3	1	2016	27.88	0.99	N/A
	2009	CORE3	3	2015	24.71	0.92	-7.00
	1999	CORE4	7	2016	4.17	0.96	N/A
	2001	CORE3	9	2006	10.11	0.98	-3.93
	2006	CORE3	12	2016	5.95	0.89	N/A
	2010	CORE3	13	2015	24.80	0.93	-4.00
	2009	CORE3	16	2015	15.14	0.97	-5.00
	2007	CORE3	18	2016	4.18	0.54	N/A
	2008	CORE4	24	2015	8.23	0.95	-13.00
	2004	CORE4	27	2016	4.41	0.87	N/A
	1996	CORE4	30	2000	3.60	0.48	0.01
	1996	CORE7	32	2016	2.31	0.84	N/A
	2002	CORE3	34	2015	2.74	0.84	-22.00
	2002	CORE3	35	2016	2.93	0.94	N/A
	2002	CORE2	39	2008	7.89	0.99	-2.10
	1996	CORE8	41	2012	1.60	0.82	-0.50
	2011	CORE3	43	2015	16.80	0.98	-16.00
	2008	CORE3	44	2016	8.67	0.83	N/A
	2012	CORE3	46	2016	16.80	0.99	N/A
	2007	CORE3	47	2015	7.08	0.96	-2.00
	1997	CORE4	48	2014	1.04	0.49	-9.00
	2006	CORE3	50	2016	4.99	0.94	N/A
Practice-Oriented Journals (PR)	2009	PR2	11	2016	18.60	0.89	N/A
	2007	PR2	14	2013	9.57	0.74	-18.30
	2007	PR2	33	2013	7.46	0.88	-6.80
	2006	PR5	40	2015	4.65	0.63	-17.00

Computer Science-Oriented Journals (CS)	2000	CS1	5	2007	9.20	0.97	-5.27
	2009	CS1	8	2016	26.75	0.87	N/A
	2004	CS1	20	2013	6.29	0.92	-5.50
	2006	CS1	22	2011	9.77	0.95	-7.60
	2000	CS1	29	2012	2.97	0.89	-3.30
	2005	CS1	31	2015	5.13	0.98	-16.00
	2007	CS1	36	2014	6.77	0.98	-2.00
	1997	CS1	37	2014	1.73	0.74	-1.50
Management-Oriented Journals (M)	2001	M1	2	2011	14.85	0.96	-9.91
	2010	M1	10	2016	21.71	0.96	N/A
	2007	M1	19	2012	11.37	0.87	-3.30
	2007	M1	21	2013	10.14	0.94	-5.70
	2005	M1	23	2013	5.81	0.60	-5.80
	2006	M1	49	2011	8.90	0.74	-2.17
Informetrics Journals (INFO)	2004	INFOM2	4	2016	8.86	0.96	N/A
	2006	INFOM2	6	2013	12.18	0.89	-8.00
	2010	INFOM2	26	2016	14.61	0.93	N/A
	2005	INFOM2	28	2014	3.65	0.64	-3.00
	2006	INFOM2	38	2009	14.50	0.99	-3.99
	2005	INFOM2	42	2015	3.34	0.84	-19.00
	2001	INFOM2	45	2010	3.10	0.94	-0.04
Other LIS Journals (OTHER)	2002	OTHER2	15	2016	5.15	0.94	N/A
	2004	OTHER2	17	2016	6.31	0.90	N/A
	2003	OTHER2	25	2015	3.57	0.88	-3.00

Table 4 depicts the average values of resulting data in terms of subject category. The subject category of CORE contains both the lowest publication year and the highest publication year. This is a possible indication that most-cited papers in LIS Core journal category may have longer citable papers compared to journals published in other subject categories, such as PR and M. Regarding the total citation rank, the average value ranges from 19 (OTHER) to 28.1(CORE). Since the subject category of OTHER consists of diverse types of journals, the ranks of a specific journal type should be considered in revealing the characteristics of the subject category. The averaged Peak Year of journal categories ranges from 2012.7 to 2015.7, while the overall averaged Peak Year is 2013. The average Peak Year of all subject categories suggests that the yearly citation would peak in the latter years of the citation window period.

Furthermore, the averaged slope tends to differ in respect to subject categories. The data provided in Table 4 suggest that the yearly citation in all subject categories would generally grow along the linear regression line. The subject category of M has the highest averaged Leftside-Slope value (12.13) compared to the averaged Leftside-Slope of other subject categories. This is a possible indication that the most-cited papers in category M are likely to receive a higher rate of citations in coming years compared to other categories. However, generalizability based on the subject categories are limited due to the small number of journals in each category.

<Tab. 4> Six Subject Categories and Average Value of Various Variables

Subject Category	Total Citation Rank	Publication Year	Peak Year	Leftside-Slope	Leftside-RSQ	Rightside-Slope
LIS Core Journals (CORE)	28.1	2004.3	2013.9	9.37	0.87	-7.04
Practice-Oriented Journals (PR)	24.5	2007.3	2014.3	10.07	0.78	-14.03
Computer cience-Oriented Journals (CS)	23.5	2003.5	2012.8	8.58	0.91	-5.88
Management-Oriented Journals (M)	20.7	2006	2012.7	12.13	0.85	-5.38
Informetrics Journals (INFOM)	27	2005.3	2013.3	8.60	0.89	-6.80
Other LIS Journals (OTHER)	19	2003	2015.7	5.01	0.91	-3.00
Overall Average	23.8	2004.9	2013.7	8.96	0.87	-7.02

The averaged total citation rank of all subject categories is lower than 19. Although OTHER subject category ranks higher than other subject categories in terms of averaged total citation count, the averaged Leftside-Slope of OTHER is lower than other categories, which indicates that the rate of receiving citation is the lowest among subject categories. The averaged leftside-RSQ of all subject categories is .87, indicating that the individual citation increases moderately along the slope line as an entire LIS journal level. Overall, the result suggests that moderate differences among subject categories can be expected in terms of publication years, total citation rank, Peak Year, and slope.

5. Correlations Among Major Variables

Several variables related to yearly citation count were identified earlier. However, to construct a formula for YCGR, it is useful to determine the major variables related to yearly citation count and find the relationship among these variables. The major variables related that can be used to formulate YCGR are as follows: Publication Year, Total Citation Count, Peak Year, Leftside-Slope, and Rightside-Slope. The relationship among these major variables related to yearly citation counts can be empirically tested using the Pearson correlation coefficient. Table 5 shows a varying degree of Pearson R correlation measurements among the variables related to YCGR. As shown in this table, there are several cells where the values are closer to 0. This indicates that there is no relationship between the two variables. Furthermore, there is a weak negative correlation between the Total Citation Count and Publication Year.

The cells in which two variables have a moderate or strong relationship are indicated with shaded color, and they are significantly different from 0 with a significance level alpha is equal to 0.05. There is a positive correlation between Total Citation Count and the Leftside-Slope (r=.58). Accordingly, as the paper receives more yearly citations, the slope becomes steeper and correlates inversely to Total Citation Rank. At the same time, there is a moderate positive correlation between the Publication Year and the Leftside-Slope (r=.66). This indicates that as the paper becomes older, a citation would increase moderately along the linear regression line.

<Tab. 5> Pearson R Correlations Among the Five Major Variables

	Publication Year	Total Citation Count	Peak Year	Leftside-Slope	Rightside-Slope
Publication Year	1.00				
Total Citation Count	0.07	1.00			
Peak Year	0.43	0.04	1.00		
Leftside-Slope	0.66	0.58	0.15	1.00	
Rightside-Slope	-0.29	-0.02	-0.43	0.02	1.00

Note: Cell values shown in shaded color are significantly different from 0 with a significance level (α=.05).

6. Rankings Based on the Five Major Variables and YCGR

The method suggested in this paper incorporates individual ranks of major variables related to the yearly citation count. YCGR is an averaged value of the five ranking measurements:

1) Total Citation Rank,

2) Leftside-Slope Rank,

3) Publication Year Rank,

4) Peak Year Rank, and

5) Rightside-Slope Rank.

The ranks of the five major variables are equally important since the average rank of all five major variables is used to determine YCGR Rank. For example, in Table 6, the average rank of 1-CORE3 is 3.2 based on the 5 individual indicators. Since this value is the lowest among all journals in the dataset, YCGR Rank is also #1. The ranks of 5 variables for the journal CORE3 are the following: Tot. Citation Rank (1), Leftside-Slope Rank (1), Pub. Year Rank (12), Peak Year Rank (1), and Rightside-Slope Rank (1). This is a paper entitled 'The Link-Prediction Problem for Social Networks' written by Liben-Nowell and Kleinberg and was published in 2007. In contrast, for the journal M1 (ranked #2 based on total citation count), the YCGR Rank of this journal is #29, having the YCGR value of 27.8.

The ranks of the five major variables are determined based on intuitive reasoning and realistic assumptions. First, the rank of the Publication Year is determined based assumption that newer publication, in general, would receive more visibility in the coming years. Thus a newer publication year is ranked higher than an older publication year. Secondly, regarding the Peak Year, the newer Peak Year is ranked higher than the older Peak Year since newer Peak Year, in general, would signify the likelihood of an inclining trend. Thirdly, regarding the Total Citation Rank, the papers receiving higher citations are ranked higher than papers receiving fewer citations. Ranking based on the total citations is typically done in this manner. Fourthly, regarding the Leftside-Slope, the steeper Leftside-Slope is ranked higher than the gradual Leftside-Slope. Finally, regarding the Rightside-Slope, gradual Rightside-Slope is ranked higher than the steeper Rightside-Slope since the steeper Rightside-Slope may indicate the declining rate of receiving citations. However, as already shown in Table 3, since the Rightside-Slope is labeled with "N/A" for the Peak Year due to not having declining citation year, a considerable number of papers have the ranking of "1" for the Peak Year rank, and for the Rightside-Slope rank.

In regards to Year-Ending, the yearly citation trend can be largely divided into two types: a) The Peak Year that ends at the Year-Ending and b) the Peak Year that ends prior to the Year-Ending. To distinguish these two groups, the Peak Year that ends at the Year-Ending is highlighted with a shaded color. Here, 17 out of 50 papers, which is approximately 1/3 of the ranked papers, show the Peak Year that ends at the Year-Ending. We may also reproduce a new list based on the shaded color: a) Peak Year rank and Rightside-Slope rank having the rank #1 (shaded color), and b) Peak Year rank and Rightside-Slope rank other than #1 (non-shaded color). Nonetheless, in contrast to ranks solely based on total citation counts, the values of all five major variables are calculated as a composite value – average rank value. Then, based on the average rank value, YCGR Rank needs to be determined. Here, the paper having YCGR Rank #1 has the lowest average rank value, whereas the paper having YCGR Rank #50 has the highest average rank value. Although Table 6 is currently sorted based on Total Citation Rank, Table 6 can be easily re-sorted according the YCGR Rank since the values of YCGR Rank has been obtained.

<Tab. 6> The Ranks of the Top 50 Cited Papers Based on Various Ranking Measurements

Paper Index	(1) Total Citation Rank	(2) Leftside-Slope Rank	(3) Pub. Year Rank	(4) Peak Year Rank	(5) Right-side-Slope Rank	Avg. Rank Value (YCGR)	YCGR Rank
1-CORE3	1	1	12	1	1	3.2	1
2-M1	2	10	40	44	43	27.8	29
3-CORE3	3	4	6	18	39	14	11
4-INFOM2	4	21	31	1	1	11.6	7
5-CS1	5	19	43	48	34	29.8	37
6-INFOM2	6	13	20	33	41	22.6	20
7-CORE4	7	38	46	1	1	18.6	16
8-CS1	8	2	6	1	1	3.6	2
9-CORE3	9	16	40	49	30	28.8	32
10-M1	10	5	3	1	1	4	3
11-PR2	11	6	6	1	1	5	4
12-CORE3	12	30	20	1	1	12.8	8
13-CORE3	13	3	3	18	32	13.8	10
14-PR2	14	18	12	33	48	25	26
15-OTHER2	15	32	36	1	1	17	15
16-CORE3	16	9	6	18	33	16.4	14
17-OTHER2	17	28	31	1	1	15.6	13
18-CORE3	18	37	12	1	1	13.8	9
19-M1	19	14	12	39	28	22.4	19
20-CS1	20	29	31	33	35	29.6	35
21-M1	21	15	12	33	36	23.4	22
22-CS1	22	17	20	42	40	28.2	30
23-M1	23	31	27	33	37	30.2	38
24-CORE4	24	23	10	18	44	23.8	24
25-OTHER2	25	41	35	18	26	29	33
26-INFOM2	26	11	5	1	1	8.8	5
27-CORE4	27	36	31	1	1	19.2	17
28-INFOM2	28	39	27	29	26	29.8	36

29-CS1	29	44	43	39	28	36.6	45
30-CORE4	30	40	48	50	18	37.2	47
31-CS1	31	33	27	18	45	30.8	39
32-CORE7	32	47	48	1	1	25.8	28
33-PR2	33	25	12	33	38	28.2	30
34-CORE3	34	46	36	18	50	36.8	46
35-CORE3	35	45	36	1	1	23.6	23
36-CS1	36	27	12	29	22	25.2	27
37-CS1	37	48	47	29	21	36.4	44
38-INFOM2	38	12	20	46	31	29.4	34
39-CORE2	39	24	36	47	24	34	42
40-PR5	40	35	20	18	47	32	41
41-CORE8	41	49	50	39	20	39.8	49
42-INFOM2	42	42	27	18	49	35.6	43
43-CORE3	43	7	2	18	45	23	21
44-CORE3	44	22	10	1	1	15.6	12
45-INFOM2	45	43	40	45	19	38.4	48
46-CORE3	46	8	1	1	1	11.4	6
47-CORE3	47	26	12	18	22	25	25
48-CORE4	48	50	47	29	42	43.2	50
49-M1	49	20	20	42	25	31.2	40
50-CORE3	50	34	20	1	1	21.2	18

Note: Cell values shown in shaded color indicate the Peak Year that ends at the Year-Ending.

Ⅴ. Discussion

Two important characteristics of individual variables of YCGR are worth-noting. First, this paper has shown varying degrees of correlation among the variables related to total citation counts. There was a moderate positive relationship between the Total Citation Count and Leftside-Slope, and between the Publication Year and Leftside-Slope. The correlations indicate that the rate of receiving citations, in general, is greater for papers that receive more overall citations and the papers that were published in more recent years. Secondly, YCGR is a composite measure that incorporates citation dependent and non-citation dependent variables. The values of Total Citation Rank, Leftside-Slope Rank, and Rightside-Slope Rank are dependent upon the citation count. More specifically, the values of Leftside-Slope Rank, Peak Year Rank, and Rightside-Slope Rank would depend on the yearly citation count, while the Total-Citation Rank, as the name suggests, would depend only on the total citation count. In contrast, it has been assessed that Publication Year Rank is a non-citation dependent variable.

YCGR Rank can be regarded as a composite measure of all related variables. More specifically, YCGR Rank is intended to reflect the likelihood of receiving more citations in the years to come by utilizing the yearly citation counts and Publication Date, which are non-citation dependent variables and the Peak Year. Because YCGR takes an average of all related variable, calculating the additional statistical characteristics among citation related variables, including YCGR Rank, can be additionally useful in order to discover distinctive patterns of the most-cited papers.

In using YCGR Rank for practical applications, the potential effects of dataset size and procedure in creating a dataset must be taken into consideration. For example, altering the Year-Starting and citation window period may cause some papers to fallout from the sampled dataset. In contrast, increasing the starting publication period and Year-Starting of the citation window period would result in a greater pool of population data. In the case of citation period, the result may also have been influenced by incomplete data since the cut-off year was 2016, which was two years ago. In terms of sample size, the cut-off number of 50 was used to extract only the top 50. The quantity of 50 was somewhat arbitrarily chosen as standard. Since changing the cut-off number of articles would likely change the rank based on YCGR, more empirical examination is needed to understand the effects of using varying cut-off points.

By comparing the result of this study to the earlier study, we can notice the prominent journals that published the most-cited papers. In this study, 42% of the most-cited papers were published by *JASIST* and *Journal of Documentation.* In contrast, the result conducted by Blessinger and Hrycaj (2010) showed that 62% of the most-cited papers were published in these two journals. In comparing the two studies, however, we need to consider the fact that the publication years and the journals used in creating a dataset are different from each other. To compare the result of one study to another, a more standardized journal based dataset needs to be used. Besides the issue of using differing datasets, the journal size can become an impeding factor in comparing the results of this study to the earlier studies. Since an unequal number of journals were used, a standardized set of journals is desirable in comparing the result of differing studies.

Nonetheless, the YCGR based ranking method suggested in this paper has an implication on the current features offered by the bibliographic databases. From the perspectives of researchers, developing a strategy to select influential papers would be much more convenient if the criteria for the variables can be set and customized through using the current bibliographic databases. Because changing the values of variables may produce a different set of papers, researchers may want to rank the papers by various formulas that could be offered as a part of the bibliographic database feature. Multiple views offered by a bibliographic database would allow a user to view the trend based on the yearly citation counts from multiple perspectives. In the case where a user decides to view a ranked list based on a YCGR Rank, the total citation rank would be shuffled and appear in a non-sequential order. This can be an alternative means to highlight papers according to the likelihood of receiving future citations. Such a feature would allow users to generate various ranked lists. Consequently, it could be a valuable means of identifying some niche topical areas of research.

Ⅵ. Conclusion

The main purpose of this work is to demonstrate various ways to view the most-cited articles based on YCGR. YCGR can be obtained by taking the average rankings of various variables related to yearly citations: Total Citation Rank, Leftside-Slope Rank, Publication Year Rank, Peak Year Rank, and Rightside-Slope Rank. Considering the ranking of all variables related to the citation count, the average value should indicate the likelihood of receiving future citations. The approach described in this paper is rather straightforward. This paper showed that YCGR can be obtained by taking related variables into account. The result of the paper suggests that the YCGR Rank could be used as an alternative method to view some notable papers in a research domain. Researchers who are interested in exploring various growing trends of yearly citation counts may employ the method described in this paper. It should be an effective method to view the most-cited papers from the perspective of receiving a higher rate of citation counts in the coming years.

This paper used an analytical approach to demonstrate the utility of YCGR. Since generalizing the characteristics of YCGR was limited due to using only one dataset, a much larger pool of LIS journals with varying publication years should be considered in order to find general patterns associated with the YCGR Rank. To reinforce this paper’s suggested method, additional empirical work would be useful in order to reveal the general characteristics of most-cited papers in LIS. This could ultimately lead to the development of variants of the paper’s proposed method that researchers can use to assess the key papers in a domain of interest.

References

Aksnes, Dag W. 2005. *Citations and Their Use as Indicators in Science Policy: Studies of Validity and Applicability Issues with a Particular Focus on Highly Cited Papers.* thesis, University of Twente. Enschede-West, Netherlands.

Aversa, Elizabeth. 1985. "Citation Patterns of Highly Cited Papers and Their Relationship to Literature Aging: A Study of the Working Literature." *Scientometrics,* 73(6): 383-389.

Barabási, Albert-László, Chaoming Song, and Dashun Wang. 2012. "Handful of Papers Dominates Citation." *Nature,* 491(7422): 40-41.

Bauer, Johann, Loet Leydesdorff and Lutz Bornmann. 2016. "Highly Cited Papers in Library and Information Science (LIS): Authors, Institutions, and Network Structures." *Journal of the Association for Information Science and Technology,* 67(12): 3095-3100.

Blessinger, Kelly and Paul Hrycaj. 2010. "Highly Cited Articles in Library and Information Science: An Analysis of Content and Authorship Trends." *Library & Information Science Research,* 32(2): 156-162.

Brinjikji, Waleed, Alexa Klunder and David F. Kallmes. 2013. "The 100 Most-Cited Articles in the Imaging Literature." *Radiology,* 269(1): 272-276.

Davis, Fred Davis. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS quarterly,* 13(3): 319-340.

Davis, Philip M. and Angela Cochran. 2015. "Cited Half-Life of the Journal Literature." arXiv preprint arXiv:1504.07479.

Garfield, Eugene. 1987. "100 Citation Classics from the Journal of the American Medical Association." *JAMA,* 257(1): 52-59.

Ivanović, Dragan. and Ho, Yuh-Shan. 2016. "Highly Cited Articles in the Information Science and Library Science Category in Social Science Citation Index: A Bibliometric Analysis." *Journal of Librarianship and Information Science,* 48(1): 36-46.

Jafarzadeh, Hamid, Alireza Sarraf Shirazi and Lars Anderson. 2015. "The Most-Cited Articles in Dental, Oral, and Maxillofacial Traumatology During 64 Years." *Dental Traumatology,* 31(5): 350-360.

Feijoo, Javier F. et. 2014. "The 100 Most-cited Articles in Dentistry." *Clinical Oral Investigations,* 18(3): 699-706.

Lee, Jong-Wook and Hye-Rin Bak. 2016. "Characteristics of Korean Researchers Through Bibliometric Analysis of Papers Published in International LIS Journals." *Journal of Korean Library and Information Science Society,* 47(1): 217-242.

Levitt, Jonathan and Mike Thelwall. 2008. "The Most Highly Cited Library and Information Science Articles: Interdisciplinarity, First Authors and Citation Patterns." *Scientometrics,* 78(1): 45-67.

Lokker, Cynthia et al. 2008. "Prediction of Citation Counts for Clinical Articles at Two Years Using Data Available Within Three Weeks of Publication: Retrospective Cohort Study." *BMJ,* 336(7645): 655-657.

Newman, M. E. J. 2014. "Prediction of Highly Cited Papers." *EPL (Europhysics Letters),* 105(2): 28002.

Patience, Gregory S. et al. 2017. "Citation Analysis of Scientific Categories." *Heliyon,* 3(5): e00300.

Walters, Glenn D. 2011. "The Citation Life Cycle of Articles Published in 13 American Psychological Association Journals: A 25-Year Longitudinal Analysis." *Journal of the Association for Information Science and Technology,* 62(8): 1629-1636.

Walters, William H. and Esther Isabelle Wilder. 2015. "Worldwide Contributors to the Literature of Library and Information Science: Top Authors, 2007–2012." *Scientometrics,* 103(1): 301-327.

Yan, Rui et al. 2001. "Citation Count Prediction: Learning to Estimate Future Citations for Literature." *Proceedings of The 20th ACM International Conference on Information and Knowledge Management.* ACM:1247-1252.