International Journal of Information Science and Management
Vol. 18, No. 1, 2020, 25-33

Tracing the Origin of Information Seeking Behavior by Reference Publication Year Spectroscopy (RPYS): Scientific Publication Based on ISC Database

Zohre Eftekhar
PhD in Knowledge and Information Science,
Shiraz, Iran
Zohre_eftekhar@yahoo.com

Abstract

This paper aims to introduce the quantitative method named Reference Publication Year Spectroscopy (RPYS). With this method one can determine the historical roots of research fields and quantify their impact on current research. Using a quantitative method named reference publication year spectroscopy (RPYS), this research tried to determine the historical roots of information seeking behavior research. RPYS pave the way for determining the significant years and works in information seeking behavior. The main problem is it can reveal whether Information Seeking Behavior has the same or different impact on these journals in terms of citation. Using scientometrics method, the initial data of this study have been extracted based on ISC Database. The revised data were analyzed and visualized in Excel. The distribution of cited references in information seeking behavior revealed four peaks within 20th century. Moreover, our analysis identified 4 peaks between 1989 and 2000 in the field of information seeking behavior, occurred respectively in 1989, 1993, 1998, and 2000. Based on the study findings, it seems that information seeking behavior research has been shaped intellectually by fields such as library and information science, management, etc. Moreover, it has been influenced by some theories and theoretical works.

Keyword: Reference Publication Year Spectroscopy, RPYS, Information Seeking Behavior, Information Behavior, Scientometrics, Islamic World Science Citation Center, ISC.

Introduction

In the current atmosphere of research evaluation on all levels of scientific activities (single researchers, research groups, institutions, and countries), bibliometrics can be considered as an instrument for evaluation purposes (Pendlebury, 2008). For example, bibliometric data is one of the most important data sources for university rankings: The Leiden Ranking (Waltman et al., 2012) is solely based on bibliometric data and for the Academic Ranking of World Universities (http://www.shanghairanking.com) and the Times Higher Education World University Rankings (http://www.timeshighereducation.co.uk/world-university-rankings) this data plays an important role. The focus on evaluation purposes in the use of bibliometrics may close one's eyes to other possibilities of using this data. A good example for another use is the VOSViewer tool (http://www.vosviewer.com), which allows visualizations of scientific

activities beyond evaluation (citation relations among scientific fields, university profiles and collaborations, co-citations of journals, etc.) (Bornmann, Thor, Marx, and Leydesdorff, 2016).

Bornmann and Marx (2013) argue for broadening the perspective in bibliometrics by complementing the (standard) times cited with cited reference analyses using field-specific impact measurements. For such a cited reference analysis, they propose to extract all cited references from a field-specific publication set and to analyze which papers (scientists or journals) have been cited most often and in which years. The main problem is whether Information Seeking Behavior has the same or different impact on these journals in terms of citation

(Eftekhar, & Hayati, 2016) (Eftekhar, & Ziaei, 2016). RPYS is based on the analysis of the frequency with which references are cited in the publications of a specific research field in terms of the publication years of these cited references. The origins show up in the form of more or less pronounced peaks mostly caused by individual publications that are cited particularly frequently (Bornmann, et al., 2016).

Reference Publication Year Spectroscopy (RPYS)

In a previous publication, rather than starting with citations of the publications in a specific research field, for certain issues we proposed reversing the perspective and analyzing the papers referenced in the publications in the same research field in order to determine the impact of publications, authors, institutions or journals within that field (Bornmann & Marx, 2013). A cited reference analysis with specific emphasis on the publication years of the references can be used to quantify the significance of historical publications and to reveal the historical roots of a given research field. The analysis of the publication years of references is not a new bibliometric approach but has already been discussed by Price (1970). And it was for instance applied by Van Raan to measure the growth of science and to detect important breakthroughs in science without pre-defining any fields (Marx & Bornmann, 2014).

Quantitative analysis of the reference publication years (RPYs, not to confuse with the method RPYS) in a specific research field shows that frequently occurring RPYs become more differentiated towards the past and mostly show up as distinct peaks. If one analyzes the publications underlying these peaks, one will find that during the 19th and the first half of the 20th century they are predominantly formed by single relatively highly cited publications. These few, particularly frequently cited publications contain as a rule the historic papers (or even the historical roots) most relevant for the evolution of a specific research field which should be taken into consideration when discussing its history. Their specific role can only be revealed by careful analysis through experts in the relevant field (ibid).

Recently, RPYS has been used to examine the historical roots in some research fields: Marx, Bornmann, Barth, & Leydesdorff (2014) used research on graphene and on solar cells to illustrate how RPYS functions. Leydesdorff, Bornmann, Marx, and Milojevic (2014) investigated the historical origins of iMetrics (information metrics, bibliometrics, and scientometrics) in scholarly literature. For example, they found that Lotka (1926) can be considered as the first source, but the intellectual program of iMetrics was especially shaped in the early 1960s. Whereas Barth, Marx, Bornmann, and Mutz (2014) examined the origins of the Higgs boson research and combined RPYS with a segmented regression analysis, Wray and Bornmann (2015) took a closer look at the roots of the philosophy of science. As the results of Marx and Bornmann (2014) show, RPYS can not only be applied to the identification of origins,

but also to reveal scientific legends: “Charles Darwin, the originator of evolutionary theory, was given credit for finches he did not see and for observations and insights about the finches he never made” (p. 839). The analysis validated bibliometrically the known fact that a book from 1947 is the origin of the term “Darwin finches” (Lack, 1947; Bornmann, et al., 2016).

Most RPYS papers are based on software which can be downloaded at http://www.leydesdorff.net/software/rpys. Recently, Comins and Hussey (2015) used RPYS to investigate the impact of Viterbi algorithm first published by Andrew Viterbi in 1967. They extended the method of RPYS with heat maps with the goal of comparing the results of different RPYS (Bornmann, et al., 2016).

Methods

The research method used in this study is a Quantitative method named Reference Publication Year Spectroscopy (RPYS). With this method one can determine the historical roots of research fields and quantify their impact on current research. Using a quantitative method named reference publication year spectroscopy (RPYS), this research tried to determine the historical roots of information seeking behavior research. RPYS pave the way for determining the significant years and works in information seeking behavior. Using scientometrics method, the initial data of this study have been extracted based on ISC Database at the end of 2016. The first step in RPYS is to select the publications for a certain research field and extract all the references from them. The second step is to establish the distribution of the frequencies of the cited references over the RPYS and from this determine the early RPYS cited most frequently. The third is to analyze these RPYS for frequently cited historical publications.

The publications dealing with "Information Seeking Behavior" were selected in the ISC database by searching with these terms: "information behavior", "information need", "information seek", "information use", "information search behavior". All cited references with reference publication years limiting the reference publication years to the time period 1958 to 2014 have been selected from the references of the complete set of 748 papers on "Information Seeking Behavior" published within scientific literature since 1958. The revised data were analyzed and visualized in Excel.

Findings

The distribution of cited references in information seeking behavior revealed four peaks within 20th century. Moreover, our analysis identified four peaks between 1989 and 2000 in the field of information seeking behavior. Figure 1 shows the distribution of the cited references limiting the reference publication years to the time period 1958 to 1989. Figure 2 shows the distribution of the cited references limiting the reference publication years to the time period 1990 to 1994. Figure 3 shows the distribution of the cited references limiting the reference publication years to the time period 1995 to 1999. Figure 4 shows the distribution of the cited references limiting the reference publication years to the time period 2000 to 2014.

Figure 1 shows the distribution of the cited references limiting the reference publication years to the time period 1958 to 1989. However, a large number of references with early reference publication years usually prove to be erroneous.

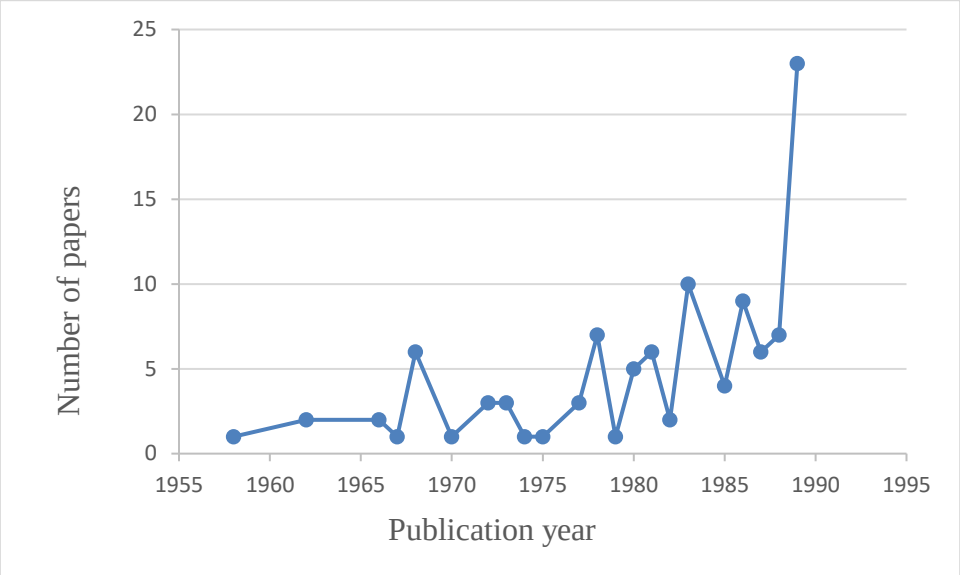

Figure 1. shows the distribution of the cited references limiting the reference publication years to the time period 1958 to 1989

The most clearly pronounced peak in Figure 1 can be assigned to the reference publication year 1989 with 23 out of 104 citations among 1958 to 1989. Most citations refer to article by D. Ellis entitled “A behavioral approach to information retrieval design" published in *Journal of Documentation* in 1989. And an article by J. Bichteler & D. Ward entitled “Information seeking behavior of geoscientist" published in *Special Library* in 1989. The former is about Ellis's behavioral model of information search strategies. Ellis's information seeking behavior model included six elements that he called stages. These stages are: starting, chaining, browsing, differentiating, monitoring and extracting. Most citations refer to this article after Years. However, a careful reading of these articles shows that they are the origin of information seeking behavior.

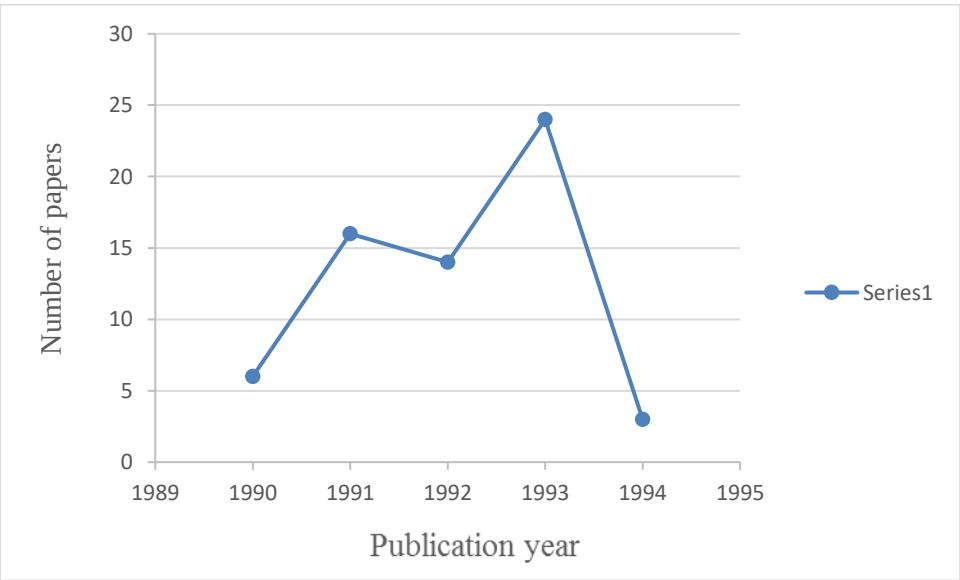

Figure 2: the distribution of the cited references limiting the reference publication years to the time period 1990 to 1994

The most clearly peak in Figure 2 can be assigned to the reference publication year 1993 with 24 out of 63 citations among 1990 to 1994. Most citations refer to article by D. Ellis, D. Cox & K. Hall entitled “A comparison of the information seeking patterns of researchers in the physical and social sciences " published in *Journal of Documentation* in 1993. There is no basic and fundamental difference between Ellis new model proposed in this article with the previous one except he added two new features in the new research. This features brought about to achieve more citations positioned him at the peak.

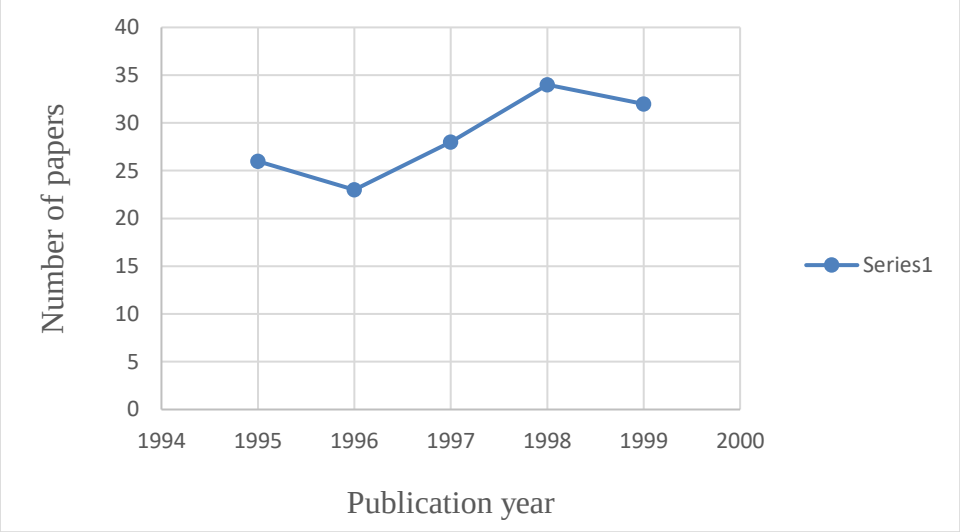

Figure 3: the distribution of the cited references limiting the reference publication years to the time period 1995 to 1999

The most clearly peak in Figure 3 can be assigned to the reference publication year 1998 with 34 out of 143 citations among 1995 to 1999. Most citations refer to article by S. Shoham entitled " Scholarly communication: A study of Israeli academic researchers" published in Journal of Librarianship and Information Science in 1998. This article states that despite extensive changes in higher education, institutions and libraries that have occurred during the previous 45 years since the interest in information gathering behavior began, patterns for obtaining information remain conservative and have resisted transformation. Professional periodicals are still the most important tools for obtaining professional information and monographs still play a major role.

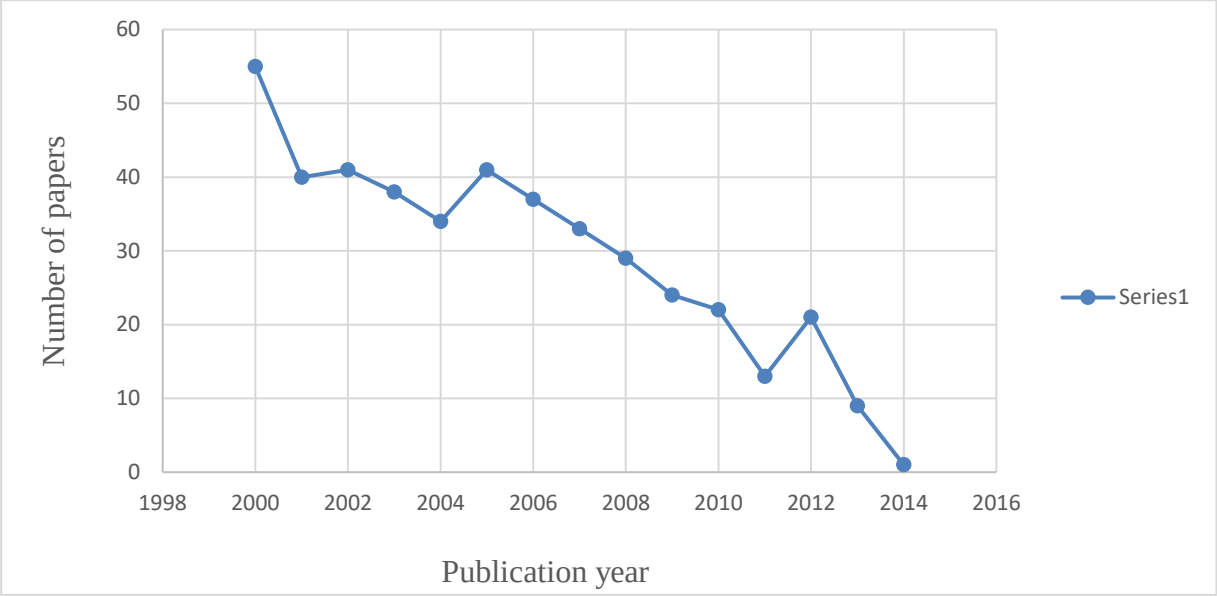

Figure 4: the distribution of the cited references limiting the reference publication years to the time period 2000 to 2014

The most clearly peak in Figure 4 can be assigned to the reference publication year 2000 with 55 out of 438 citations among 2000 to 2014. Most citations refer to article by T. D. Wilson entitled “Human information behavior " published in *Journal of Information Science* in 2000. This article is about Wilson's behavioral model. This model was called a "nested model" because Wilson pointed out that information searching behavior is a subset of information seeking behavior and that information seeking behavior is, in turn, only a subset of all possible information behavior.

Figure 5 shows four peaks between 1989 and 2000 in the field of information seeking behavior occurred respectively in 1989, 1993, 1998, and 2000.

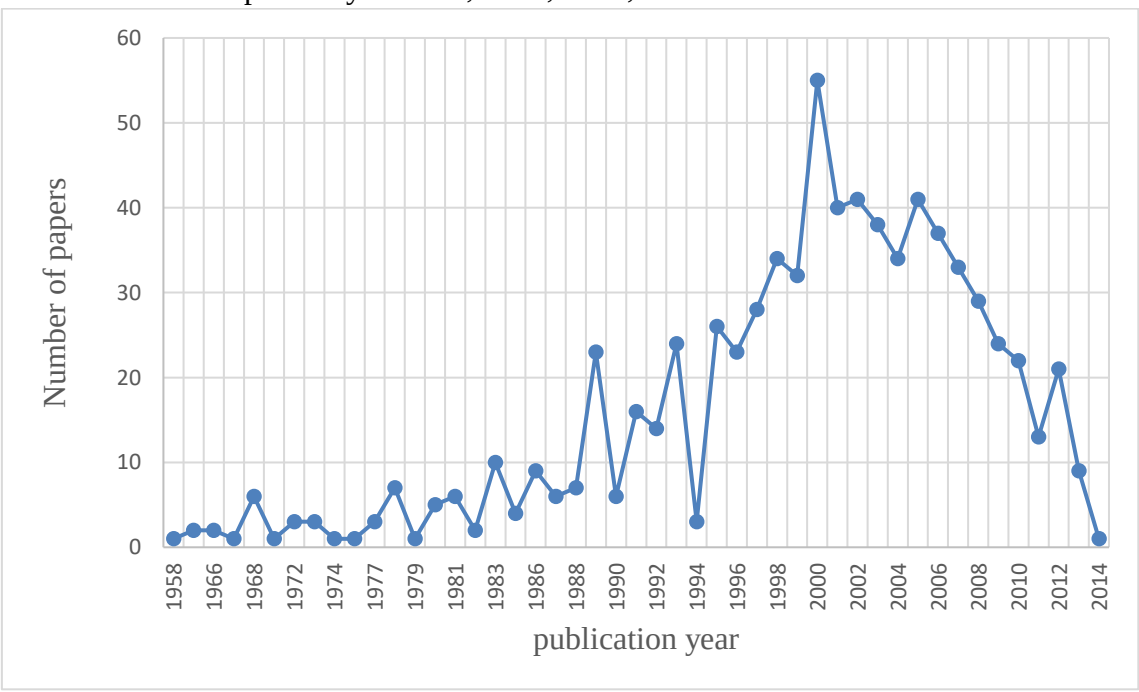

Figure 5: four peaks between 1989 to 2000

Discussion and Conclusion

Bibliometric data can be used not only for evaluation purposes, but also for scientifically related investigations in a historical context. The bibliometric cited-reference data in ISC can be traced back to the first ever published documents. RPYS was proposed by Marx, et al. (2014) to identify seminal publications in the knowledge base of a research field which are most important in a historical context. The RPYS can be used very flexibly in different contexts: (1) historical roots of research fields can be identified; (2) scientific myths can be uncovered (Marx & Bornmann, 2014); (3) most important publications for authors of specific journals can be identified (Leydesdorff, et al., 2014); (4) importance of single publications can be compared between different research fields, journals, and researchers. The origin of the term "Information Seeking Behavior" was taken as a case study to point out the potential of the RPYS method. Distribution of the cited references limiting the reference publication years to the time period 1958 to 1989 and 1990 to 1994 shown that Information seeking behaviours are derived from information need. When users feel a need for information, they fall within a process called the information seeking behaviour which is common for everyone, and includes some stages that users should follow to satisfy their information needs. The information seeking behaviour models drawn up by different researchers have different stages. Among all researchers Ellis's model is so important in those peaks. Ellis's (1989) information seeking behaviour model included six elements that he called stages. These stages are: starting, chaining, browsing, differentiating, monitoring and extracting. According to Ellis (1989), the first step in information seeking behaviours is selecting subjects and identifying information needs Process. Distribution of the cited references limiting the reference publication years to the time period 1995 to 1999 shown that there are a lot of studies on information seeking behavior of special society. Despite the extensive changes in higher learning, institutions and libraries that have occurred during the 45 years since the interest in information-gathering behavior began, including technological innovations that have brought far-reaching changes in information technology, patterns of obtaining information remain conservative and have not undergone transformations. The professional journal is still the most important tool for obtaining professional information, and the monograph still plays a major role, including for the social and natural sciences. Distribution of the cited references limiting the reference publication years to the time period 2000 to 2014 shown that Wilson's model is so important in that peak. Wilson expanded his model which has been regarded as a macro model that covered other theoretical models of information behaviour like Dervin (1983), Ellis (1989), and Kuhlthau (1993). However, Wilson (1999) presented a new model which introduced problem solving as the overall framework for information seeking, and showed that Kuhlthau's model (1991) might have conformity with various stages of the information seeking process of Ellis's (1989) model. He also maintained that Ellis's (1989) model might include a set of activities which Kuhlthau (1991, 1993) named as "Collection". In general, all information models that are mentioned in peaks above were nested within Wilson's (1997) model of information behaviour. This model was called a "nested model" because Wilson pointed out that information searching behaviour is a subset of information seeking behaviour and that information seeking behaviour is, in turn, only a subset of all possible information behavior. Generally, based on the study findings, it seems that information seeking behavior research has been shaped intellectually by fields such as library and information science, management, medical, education, psychology, quantitative

and qualitative methodologies, etc. Moreover, it has been influenced by some theories and theoretical works.

References

Barth, A., Marx, W., Bornmann, L., & Mutz, R. (2014). On the origins and the historical roots of the Higgs boson research from a bibliometric perspective. *The European Physical Journal Plus*, 129(6), 1-13. doi: 10.1140/epjp/i2014-14111-6.

Bichteler, J., & Ward, D. (1989). Information seeking behavior of geoscientist. *Special Library*, 80(3), 169-178.

Bornmann, L., & Marx, W. (2013). The proposal of a broadening of perspective in evaluative bibliometrics by complementing the times cited with a cited reference analysis. *Journal of Informetrics*, 7(1), 84–88. doi: 10.1016/j.joi.2012.09.003

Bornmann, L., Thor, A., Marx, W., & Leydesdorff, L. (2016). Identifying seminal works most important for research fields: Software for the Research Publication Year Spectroscopy (RPYS). *CollNet Journal of Scientometrics and Information Management*, 10 (1): 125-140.

Comins, J., A. & Hussey, T. W. (2015). Compressing multiple scales of impact detection by Reference Publication Year Spectroscopy. *Journal of Informetrics*, 9(3), 449-454.

Dervin, B. (1983). An Overview of sense-making research: concepts, methods and results to date. In: *International Communications Association Annual Meeting*. Dallas,Texas. Retrieved from https://faculty.washington.edu/wpratt/MEBI598/Methods/An%20Overview%20of%20Sense-Making%20Research%201983a.htm

Eftekhar, Z. & Hayati, Z. (2016). Coping with Information Resources: Identifying, Searching, Accessing, Evaluating and Using Information in Academic Libraries. *International Journal of Information Science and Management (IJISM)*, 14 (1): 1-11.

Eftekhar, Z. & Ziaei, S. (2016). Information Seeking Behavior and Information Resources Management: Mental Process of Selecting Subjects and Identifying Information Needs (Case study: Graduate Students in Women Seminaries of Shiraz, Academic Year 1393-1394). *Iranian Journal of Information Processing and Management*, 31 (4): 981-1007.

Ellis, D. (1989). A Behavioral Approach to Information Retrieval Design. *Journal of Documentation*, 45(3), 171-212.

Ellis, D., Cox, D., Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49 (4), 356-369.

Kuhlthau, C. C. (1991). Inside the search process: information seeking from the user perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.

Kuhlthau, C. C. (1993). A Principal of uncertainty for information seeking. *Journal of Documentation*, 49(4), 339–355.

Lack, D. (1947). *Darwin's finches*. Cambridge, UK: Cambridge University Press.

Leydesdorff, L., Bornmann, L., Marx, W., & Milojevic, S. (2014). Referenced Publication Years Spectroscopy applied to iMetrics: Scientometrics, Journal of Informetrics, and a relevant subset of JASIST. *Journal of Informetrics*, 8(1), 162-174. doi: DOI 10.1016/j.joi.2013.11.006.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317 - 323.

Marx, W., & Bornmann, L. (2014). Tracing the origin of a scientific legend by reference publication year spectroscopy (RPYS): The legend of the Darwin finches. *Scientometrics*, 99(3), 839-844. doi: DOI 10.1007/s11192-013-1200-8.

Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764. doi: 10.1002/asi.23089.

Pendlebury, D. A. (2008). *Using bibliometrics in evaluating research*. Philadelphia, PA, USA: Research Department, Thomson Scientific.

Price, D. J. D. S. (1970). *Smiles at the Unobtrusive*. Nature, 226(5249), 985-985.

Shoham, S. (1998). Scholarly communication: A study of Israeli academic researchers. *Journal of Librarianship and Information Science*, 30 (2), 113-121.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., Wouters, P. (2012). The Leiden Ranking 2011/2012: data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419-2432.

Wilson, T. D. (1997). Information behavior: An interdisciplinary approach. *Information Processing and Management*, 33(4), 551–572.

Wilson, T. D. (1999). Models in information behavior research. *Journal of Documentation*, 55(3), 249–270.

Wilson, T. D. (2000). Human information behavior. *Journal of Information Science*, 3(2): 49-55.

Wray, K. B., & Bornmann, L. (2015). Philosophy of science viewed through the lense of "Referenced Publication Years Spectroscopy" (RPYS). Scientometrics, 102(3), 1987-1996.