International Journal of Information Science and Management
Vol. 17, No. 2, 2019, 99-117

Identifying Information Retrieval Research Trends Using Author Co-Citation Network[1]

Hamid Alizade Zowj
Assistant Prof. Research Department of Information Management, RICeST, Iran,
halizadeh@ricest.ac.ir

Mohammad Reza Ghane
Associate Prof. Research Department of Evaluation and Resource Development, RICeST, Iran,
Corresponding Author: ghane@ricest.ac.ir

Fereshte Ehsanifar
M.A. in Scientometrics
f.ehsanifar92@gmail.com

Abstract

The aim of this study was mapping, visualizing and determining subject trends in the field of information retrieval using author co-citation network based on articles indexed in Scopus from 2005- 2018. This scientometric study was performed using co-citation analysis. Research population includes all articles indexed in Scopus in the field of information retrieval from 2005 to 2018. Therefore, 35018 papers were retrieved in this field. VOSviewer was used to analyze the author co-citation. The study indicated that a total of 604757 authors were co-cited, 212328 journals were cited. Also highly cited articles and sources were determined. Amongst countries, United States, China, United Kingdom, Germany and Canada ranked one to five, respectively. Computer science was a pioneer with regard to interdisciplinary area in IR. It is noteworthy that visualization of author co-citation in field of IR determined ten clusters, namely knowledge and information science, computer science, electronics, information retrieval, information seeking behavior, psychology, multimedia information retrieval, software engineering, ophthalmology and surgery.

Keywords: Information Retrieval, Author Co-citation Analysis, Research Trends, Visualization, Scopus.

Introduction

Scientific publications in different subject fields are expanding and research results are frequently published and distributed in the form of articles, books and other media based on the disciplines. Scientific community encounters with the huge volume of distributed information and needs to have access to the findings alongside the technological advances. Considering information needs and relevant documents the new information technologies led to appearance of a subfield, i.e. information retrieval (IR), in library and information science as an interdisciplinary field. Information, storage and effective retrieval of information seem to be absolutely essential and important to meet the needs of scholars.

The appearance of IR dated to the beginning and middle of 1950s. The term IR was coined for the first time by Mooers (1951). IR is associated with the process of representing, storing, searching and finding of information which is relevant to a requirement for information desired by a user (Ingwersen, 1992). In general, it is concerned with storage,

organization, representation and access to information based on information systems (Salton & McGill, 1983).

IR is a core research field in Information Science (Jä rvelin & vakkari, 1992). Its goal is to meet researchers' information needs accurately by focusing more on precision than recall. Therefore, to relate the community to the right information still is an important challenge in IR. The multidimensional nature of relevance motivates researchers who are interested in IR to investigate it from different approaches. Accordingly, developments are moving in different directions and trend research and identification of relevant trends regarding IR are important. In addition, IR is associated with the changes in information technologies (especially computer). The technological changes led to more studies in this field. Consequently, the growth of studies, require more investigations and determination of IR trends. Information retrieval provides a bridge among several disciplines. In this regard, library and information science and computer science are early partners. IR, as a multidisciplinary area, needs to be investigated considering its trends and future. This process provides a deep understanding of scientific fields and uncovers its connections and collaborations with relevant areas. To this end, we used scientometrics measures to analyze citations and also employed co-citation network.

Co-citation analysis has been done via different methods. Small (1973) and Marshakova (1973) considered this concept as a variable for scientific maps (Salemi & Koosha, 2013) and a method for investigating cognitive structure of scientific specialties (Ding, Chowdhury & Foo, 1999). The study of co-citation Analysis, is started with the assessment of co-cited documents and is categorized into document co-citation analysis, DCA (Small & Griffith, 1974), author co-citation analysis, ACA (White & Griffith, 1981) and journal co-citation Analysis, JCA (McCain ,1990).

We used ACA to trace the intellectual structure of IR. ACA deals with the subject relevancy among authors in a certain filed. The more co-citation of the two authors, the more subject related. ACA indicates to what extent the authors are related regarding subject connection (Borner, Chen & Boyack, 2003) as well as the intellectual structure of fields, can be mapped by a network of co-citations. When aggregated to larger sets of publications, co-citation maps indicate clusters of related scientific works (i.e., based on the same publications, as far as reflected by the cited literature). These clusters can often be identified as 'research specialties' (McCain 1990; Bayer, Smart, & McLaughlin, 1990; White & McCain 1998; Small 1999; Prime, Bassecoulard & Zitt, 2002).

The scientific structure and the overriding subject trends of IR give priority to the depiction, visualization and intellectual structure of this field, because it has much importance in storing, retrieving and facilitating information accessibility. Studies of subject trends in IR uncover the degree to which IR is an interdisciplinary area. The growth in all aspects of IR has led to cross-disciplinary studies and exponential increase in co-authorship. Hence, a periodic assessment of IR is crucial in determining the nature of its relationship with other areas. Generally, the lack of information in a given discipline may hinder an understanding of trends within profession especially interdisciplinary fields such as IR. As a result, the purpose of this study is mapping, visualizing and determining subject trends in the field of IR using author co-citation network and SCOPUS during 2005- 2018. To this end, the research questions are:

1. Who are the influential authors and what are the top journals and highly-cited articles?

2. What is the growth trend of IR?
3. What are the most productive organizations and countries?
4. What are the major subject trends in relation to IR?
5. How is the status of author co-citation analysis regarding IR?

Related works

Considering literature reviews we generally focus on information retrieval and methods regarding author co-citation.

Information retrieval

The initial scientometrics studies in the IR, are the studies of Ding et al. (1998, 1999; 2000; 2001). Their most related work is the assessment of intellectual structure in the IR field by using the ACA. They assessed the changes in intellectual foundation of IR within a two time spans 1981- 1991, 1992- 1997. The sample size includes 39 highly cited researchers from the Science Citation Index (SCI) and Social Science Citation Index (SSCI). They used the multidimensional scaling (MDS) and clustering techniques (CT) to create the two-dimensional maps to display the dynamic intellectual structure of IR based on scholars citing their works. To them ACA is not capable of mapping of intellectual structure and it is needed to get a complementary method i.e. Factor Analysis (FA) to depict the authors' research areas. IR also investigated regarding the intellectual structure (Ding, Chowdhury & Foo, 2000). Ding et al. results showed that IR field is a fully interdisciplinary field and underlies the novel fields of study. They concluded that there was a connection between IR and psychology, medical sciences and computer science. Tabatabaei and Beheshti (2007) depicted the interdisciplinary approach of IR using co-word method. They gathered data from Web of Science with 20 - years timespan. The results uncovered that the IR rooted in Library and Information Science and with the passing of time pervaded in computer science. It is noteworthy that IR has had the most cooperation with computer science and library and information science. Intellectual structure can be mapped to find the relationship among stakeholders of disciplines. Regarding this, Rorissa and Yun (2012) mapped the intellectual structure in the IR field based on the collected data from Web of science over 10 years period. They determined the highly productive authors (10 or more papers), highly cited journals, frequency and betweenness centrality of words, related fields and author's institute in the field.

Because of close relationship between IR and Information technologies, a new area, namely interactive information retrieval (IIR) came to being (Kelly and Sugimoto, 2013). Kelly and Sugimoto documented the evolution of IIR and concluded with the forward march of time the publications increased. They found two thirds of articles were published between 1997 and 2006 and co-authorship pattern based on their study was 19 percent. The citation analysis of the study showed that Salton, Belkin, Shneiderman, Saracevic, and Croft were the core authors, respectively. IR also studied from psychological and ergonomical approach. In this regard, Dinet (2014) studied the behavior of individuals concerning scio-technical system. He analyzed the behaviors of end users and according to individuals, search engine was perceived as Internet, and regarding Internet they ignored technical aspects in favor of the social networks

Historically, IR early days refered to 1950s and continued to 1960s which is a time for the foundation of the field. During these two decades Luhn's work on automatic indexing

(Williams, 2010), Cleverdon, Milis, and Keen's (1966) evaluation methodology and index experiments, Salton's (1981) "the Smart environment for retrieval system evaluation-advantages and problem areas" dealt with automatic and manual indexing and evaluation systems. Studies on retrieval models, indexing and search theories experimented vector space model and probabilistic models in 1970s and 1980s (Rocchio, 1971; Robertson, 1977; Porter, 1980). Large-scale evaluation in the light of TREC and language models developed in 1990s (Berry, Dumais & O'Brien, 1995). Studies from 2000s onward have been pointing on more applications regarding web search, machine learning, relevance feedback, and scalability (Joachims, 2002).

Author Co-Citation

There are sample studies, as Subramani, Nerur & Mahapatra (2003), Andrews (2003), Osareh & McCain (2008), Nerur, Rasheed & Natarajan (2008), Ma, Dai, Ni & Li (2009), Taherian & Osareh (2011), Zhao & Strotmann (2011a), Chen & Lien (2011) and Daniali & Naghshineh (2014) in which ACA method was used to visualize the intellectual structure of the fields, such as knowledge management, medical informatics, chemistry, strategic management, information science, stem cells and e-learning. Some of them will be alluded as follows.

ACA was used to map and analyze the intellectual structure of scientific fields (Jeong, Song, & Ding, 2014). This kind of analysis is rooted in citation networks and presents the strength and direction of citation impact. This feature paved the way in employing this method in trend patterns of intellectual structures as is evident in "information retrieval, international management, strategic management, and e-leaning" (ibid, p.198). Regarding ACA method and techniques, researchers did not reach a consensus on first-author and all-author for co-citation counting. Some researchers believe that all-author co-citation is better than first-author (Zhao and Logan, 2002). Others focusing on last-author citation counting (Zhao and Stroman, 2011a). Daniali & Naghshineh (2014) mapped the intellectual structure in the field of image retrieval during 2001- 2012 and determined highly cited authors, highly productive authors and authors with highest centrality, sigma and burst detection. They concluded that not all highly productive authors in this field necessarily obtain the above-mentioned features. Intellectual structure in information science is assessed by Ma et al. (2009) using Google Scholar. They chose 31 superior authors of information science in China and analyzed them through ACA. They also investigated the direction and trends of the domain. They showed chicness researchers paid more attention to theoretical studies than empirical ones. Stem cells also are studied during 2004-2009 (Zhao & Strotmann, 2011b) using PubMed and SCOPUS, and 200 highly cited authors regarding their co-authorship were identified. The study found despite the theoretically highly interdisciplinary nature of the field, stem cell researches have been dominated by a few central medical research areas e.g. cancer and regenerative medicine of the brain, the blood, the skin, and the heart, and a core of cell biologists trying to understand the nature and the molecular biology of stem cells along with biotechnology researchers investigating the practical identification, isolation, creation, and culturing of stem cells.

Researchers in other scientific areas such as e-learning in terms of management information systems approached the intellectual structure using ACA (Chen & Lien, 2011). They classified and analyzed the research topics in clusters, at the international level and

national one in Taiwan. They concluded that Taiwanese authors had more influence on business training studies, while international authors paid more attention to the user's mental reactions on learning context.

To sum up, there are many studies that dealt with the intellectual structure of academic disciplines. By reviewing literature, any studies are not observed in the field concerning intellectual structure in the area of IR using ACA network in SCOPUS. Some researchers, who mapped the intellectual structure in IR field, had used other methods except ACA method i.e. co-authorship and co-words (Rorissa & Yuan, 2012; Tabatabaei and Beheshti, 2007). Some studies were limited to a narrow subject area such as Ding et al. (1999) just focused on communicational structure among researchers and ignored other structural aspect e.g. journals, articles, institutions and active academic disciplines. In addition, most studies collected data from WoS and the analyses were based on the first author considering co-citation analysis. In this regards, it is perceived that the results cannot be comprehensive (Zhao and Strotmann, 2011a).

Methods and Data

We used descriptive-analytical method to illustrate the thematic similarities between authors' sources and trends in IR and analyzed the findings through scientometrics indicators and social network analysis. We used the words IR and combination of words to collect data based on the methods implemented by White & McCain (1988), Ding et al. (1999), Gmür (2003) and Rorissa & Yuan (2012). The search strategy is as follows:

TITLE-ABS-KEY ("information retrieval") AND PUBYEAR AFT 2004 AND PUBYEAR BEF 2019 AND DOCTYPE (article)

To provide a comprehensive structure of disciplines development; a more full-fledged data set is needed. To achieve this aim we extracted data from SCOPUS because it covers more Interdisciplinary fields than Web of Science (WoS) and it is also preferable to Google Scholar (Teixeira, 2011; Zhao & Strotmann, (2011a). Using WoS provides the first author for analyzing the ACA method (ibid). This function has not made this analysis possible for other co-authors in a source. This is a defect for ACA and shows the deficiency and uncertainty in the results of studies. In order to visualization ACA, the VOSviewer software is used and for doing scientometrics analysis and mapping needed diagram, Excel software is used.

Research population includes the articles on IR which used the expression "Information Retrieval" in their titles, abstracts and keywords during 2005- 2018. We extracted 35018 records in which 604757 authors were involved.

Findings

Top Authors

Core authors are effective researchers in a certain field whose thoughts and theories have the most influence on the advancement of academic disciplines. In this regard, the preeminent authors in IR were introduced from two standpoints: the highly cited authors and the authors with the most amount of scientific productions. Table 1 shows the top ten authors based on the most amount of citation they have achieved in IR.

Table 1
The ten highly cited author

Rank	Author	Citation #	Rank	Author	Citation #
1	Wang, J.	2760	6	Wang, Y.	2179
2	Salton, G.	2643	7	Li, J.	2160
3	Zhang, Y.	2611	8	Li, H.	2156
4	Croft, W.B.	2450	9	Li, X.	2080
5	Zhang, J.	2230	10	Dumais, S.T.	2058

The above table shows that *Wang, J.* from UCL with 2760 citations, is known as the highly cited author in IR in years 2005- 2018. Table 2 also displays a ranked list of the most productive authors in the IR fields with Lin, H. at the top position.

Table 2
The ten highly productive authors

Rank	Author	Article#	Rank	Author	Article #
1	Lin, H.	40	6	Rindflesch, T.C.	31
2	Demner-Fushman, D.	37	7	Liu, H.	28
3	Cimino, J.J.	36	8	Wilczynski, N.L.	27
4	Lu, Z.	36	9	Boughanem, M.	26
5	Haynes, R.B.	33	10	Rubin, D.L.	25

Top Journals

The Journals as the most frequently used media publish research findings in subject fields and index the manuscripts bibliographic information as well as full-text in subject oriented and citation databases. Our results showed that 212328 journals were cited. Considering citation counts ten top journals are shown in table 3. The Journal *Nucleic Acids Research* has received 12596 citations and it is in the first place of the highly cited IR journals. More breakdowns were shown in Table 3.

Table 3
The ten highly cited journals

Rank	Journal	Citation #	Rank	Journal	Citation #
1	*Nucleic Acids Research*	12596	6	Science	5816
2	*IEEE Transactions* on *Pattern Analysis* and Machine Intelligence	8570	7	Nature	5660
3	JASIST*	8003	8	Proc natl acad sci usa[1]	4499
4	Bioinformatrics	7423	9	*IEEE Transactions* on *Image* Processing	4298
5	Information Processing and Management	6354	10	The Journal of Neuroscience	4088

* Journal of the American society for information science and technology
[1] Proceedings of the National Academy of Sciences of the United States of America

Ten highly productive journals in the IR field based on the number of published articles are shown in Table 4.

Table 4
The ten highly productive journals

Rank	Journal	Article #	Rank	Journal	Article #
1	Nucleic Acids Research	757	6	Plos One	505
2	BMC Bioinformatics	749	7	Information Processing and Management	499
3	IEEE Transactions on Image Processing	603	8	Bioinformatics	325
4	AMIA Annual Symposium Proceedings	515	9	Expert Systems with Applications	319
5	IEEE Transactions on Pattern Analysis and Machine Intelligence	511	10	JASIST	313

The Nucleic Acids Research, with 757 articles, is in the first place of highly productive journals. Comparing table 3 and 4 we conclude that Nucleic Acids Research, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Image Processing, BMC Bioinformatics, Information Processing and Management, JASIST and Bioinformatics are included in the highly productive as well as highly cited journals. Probably the journals with the highest articles, received the most citations.

Highly cited articles

The highly cited articles in a field form the theoretical foundation of a field. These articles show the consistency and steadiness of a field by publishing the theories and standpoints of a field in continuance. Table 5 shows top five highly cited reference, which have received the most citations from IR articles during 2005- 2018. Results showed the most citations are given to the papers which are dealt with the principle of IR. Furthermore, authors of these papers are among the core authors.

Table 5
The five most cited reference by IR field

Rank	Title	Authors	Source	Year	Citation #
1	Latent dirichlet allocation	Blei,d.m., Ng,a.y., Jordan,m.i.	Journal of Machine Learning Research	2003	297
2	An algorithm for suffix stripping	Porter, m.f.,	Programe	1980	276
3	Modern information retrieval	Baeza-yates, r., Ribeiro-neto,b.	Addison-wesley	1999	233
4	Introduction to modern information retrieval	Salton, g., Mcgill, m.j.	Mcgraw-hill	1983	213
5	Introduction to information retrieval	Manning,c.d., Raghavan,p., Schutze,h.	Cambridge University Press.	2008	196

Articles published in IR during 2005- 2018 ranked based on citations. Table 6 shows the highly cited articles in IR.

Table 6
The top five highly cited articles in IR

Rank	Article Title	Authors	Source Title	Year	Citation #
1	Fast and accurate short read alignment with Burrows-Wheeler transform	Li,H., Durbin, R.	Bioinformatics	2009	13592
2	The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments	Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T.	Clinical Chemistry	2009	6401
3	Dropout: A simple way to prevent neural networks from overfitting	Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.	Journal of Machine Learning Research	2014	6371
4	Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists	Huang, D.W., Sherman, B.T., Lempicki, R.A.	Nucleic Acids Research	2009	6131
5	A performance evaluation of local descriptors	Mikolajczyk, K., Schmid,C.	IEEE Transactions on Pattern Analysis and Machine Intelligence	2005	4954

The manuscript authored by Li and Durbin has received the most citations. It is published in Bioinformatics journal which is a core journal among information retrieval journals. The most citations have been given to the articles published in 2009. Three highly cited articles in this table were published in cited journals from which, two articles were published in journals assigned to artificial Intelligence.

Growth trends of scientific productions

The growth diagram in IR scientific products showed that the year 2011 with 2073 documents had the least production and the year 2006 with 3374 documents had the most productions (figure 1).

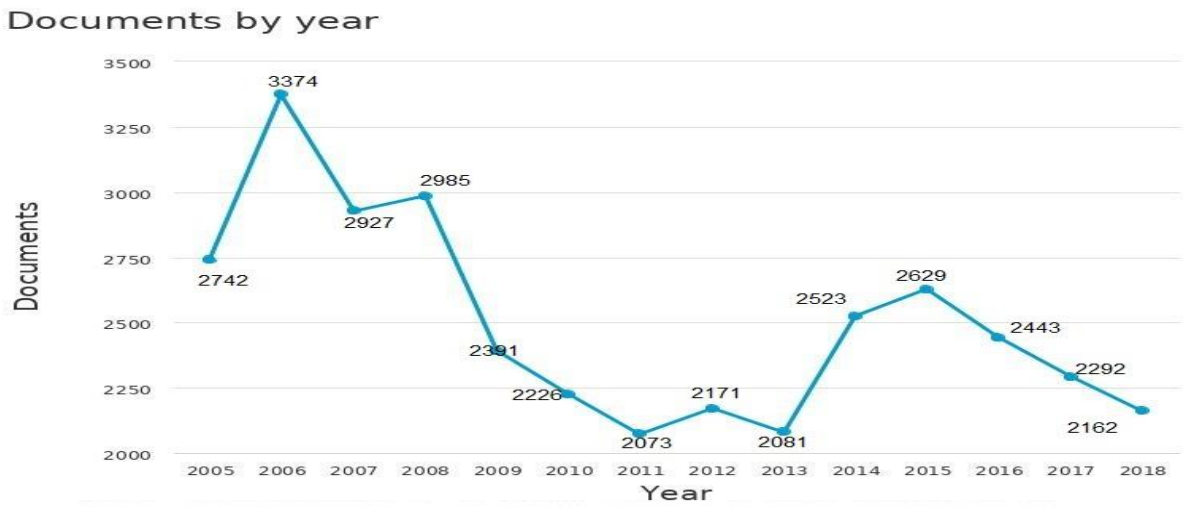

Figure 1. The growth trends of scientific production in IR (2005- 2018)

The figure 1 suggests a downtrend and fluctuation. The most productive years are 2006, 2008 and 2015. While in years 2011, 2013 and 2018 reached its least number of productions.

Active Organizations

Author affiliation is important in showing which organizations are active in a field. To this aim highly productive organizations were recognized (Table 7) using data analysis that SCOPUS has done from retrieved data sets.

Table 7
highly productive organizations in IR

Rank	Author Affiliation/ University	Article #	Rank	Author Affiliation/ University	Article #
1	Chinese Academy of Sciences	502	6	UCL (University College London)	252
2	IEEE (Institute of Electrical and Electronics Engineers)	341	7	Ministry of Education China	242
3	CNRS (Centre National de la Recherche Scientifique)	313	8	National Institution of Health, Bethesda	225
4	Tsinghua University	275	9	National Library of Medicine	222
5	University of Toronto	255	10	Zhejiang University	220

Chinese Academy of Sciences was the most productive institute in this period. IEEE is in the second place. The journals which are published by this institute are also member of the most productive and the most cited journals in field of IR.

Most productive countries

What countries are active in this area and attracted their researchers' attention to IR? Table 8 depicts the most productive countries in IR.

Table 8
Ranking ten - top countries in scientific production

Rank	Country	Article #	Rank	Country	Article #
1	United States	9925	6	France	1464
2	China	5841	7	Spain	1374
3	United Kingdom	3245	8	India	1369
4	Germany	1876	9	Australia	1361
5	Canada	1689	10	Japan	1248

Data analysis has determined 30 countries as producers of scientific production of this field. Amongst countries, United States, China, United Kingdom, Germany and Canada ranked one to five, respectively.

Highly productive subject areas

IR is one of the main subject fields of library and information science. This field aims to increase speed and accuracy in retrieval and accessibility to meet researchers' information needs. For this purpose, different systems and software are created which suggest the interdisciplinary studies of computer science and library and information science (Rorissa & Yuan, 2012). Their results showed that 82.79% of citations in IR are received by computer science and library and information science.

Table 9
subject areas based on numbers of articles in IR (2005- 2018)

Rank	Subject areas	Article #	Rank	Subject areas	Article #
1	Computer Science	17121	6	Mathematics	4287
2	Medicine	8992	7	Decision Sciences	1681
3	Engineering	7097	8	Neuroscience	1666
4	Biochemistry, Genetics and Molecular Biology	4706	9	Physics and Astronomy	1436
5	Social Sciences	4468	10	Agricultural and Biological Sciences	1222

Table 9 shows subject areas that ordered based on articles numbers. Computer science with 17121 documents is assigned as the most active discipline in IR and in that regard, medicine, engineering, and biochemistry are the highly productive fields. The results show despite the IR roots in library and information science (Tabatabaei & Beheshti, 2007), which is categorized in social sciences, other fields have been more active in advancing this field.

Common Keywords

As it is expected the most frequently used keyword is *Information Retrieval* (used in 28906 articles). Other keywords such as *Human, Article, Information Storage and Retrieval, Algorithm* are the preferred usage words. The keywords were used in 19142, 15114, 8102 and 5520 articles, respectively. Findings showed that half of the expressions were used in the field of IR articles, related to computer science and Internet which are dominant subject areas of IR.

Author Co-Citation Network

To display relationships between highly cited authors in IR, the VOSviewer software was used to map author co-citation network. ACA is defined as the frequency with which two authors are cited together by other authors. The more co-citations two authors receive, the higher their co-citation strength, and the more likely they are semantically related (Small, 1973).

VOSviewer 1.6.11 version provides two visualizations, namely the network visualization and the density visualization. In network visualization, authors are indicated by their label and, by default, also by a circle. For each item, the font size of the item's labels and the size of the item's circle depend on the weight of the item (Van Eck and Waltman, 2015). For instance, the greater nodes indicate the more citations.

Image 1 shows network visualization for 5605 co-cited authors who have received at least 40 citations from others. In this image, each circle displays one author that some other authors' names are shown on them. Every cluster is displayed by a color, in which numbers of authors are determined as nodes with different size. Bigger nodes are the authors with the most number of co-citation.

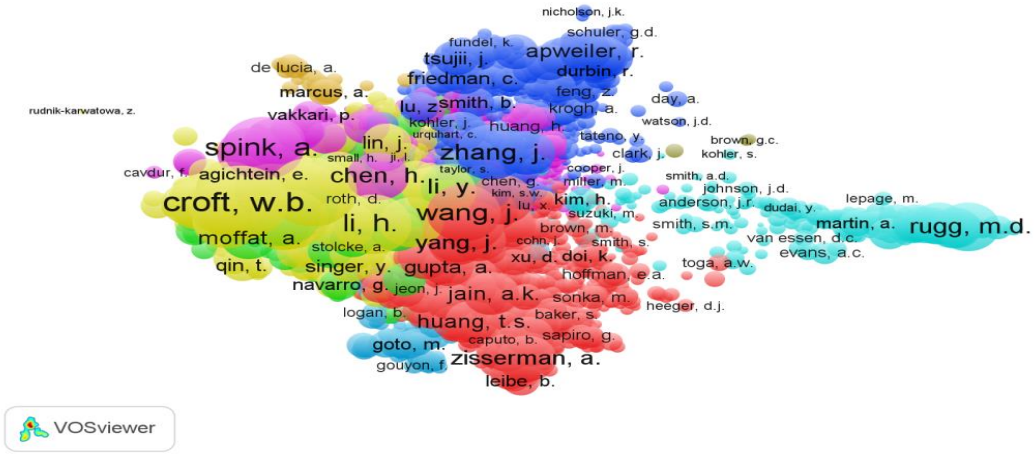

Image 1. Visualizing author co-citation network in field of IR (2005- 2018)

In IR network visualization of ACA ten clusters are determined. Every cluster is defined respectively, based on numbers of nodes they have contained in itself (Table 10). Also, the subject of each cluster is determined based on subject trend of the top author of that cluster. Therefore, the full name of every individual Google in personal web site or Google Scholar.

Table 10
Introducing top authors of each cluster and determining their subject field

Cluster	Authors	Node#	Co-citation	Citation	Subject field
1	Wang, j.	1316	70043	2760	Library and Information Science
2	Baeza-yates, r.	1256	76565	1657	Computer Science
3	Zhang, j.	1409	66346	2230	Electrical Engineering
4	Croft, w.b.	1965	148284	2450	Information Retrieval
5	Spink, a.	1248	85759	1472	Information-seeking Behavior

6	Rugg, m.d.	382	52210	1192	Psychology
7	Goto, m.	117	17271	578	Multimedia Information Retrieval
8	Marcus, a.	34	14916	470	Software Engineering
9	Sharma, s.	3	4331	147	Ophthalmology
10	Hughes, j.	2	1191	70	Surgery

Cluster one, shows the authors of library and information science. This cluster has the most nodes and authors who were cited in field of IR as the most interdisciplinary discipline (Image 2).

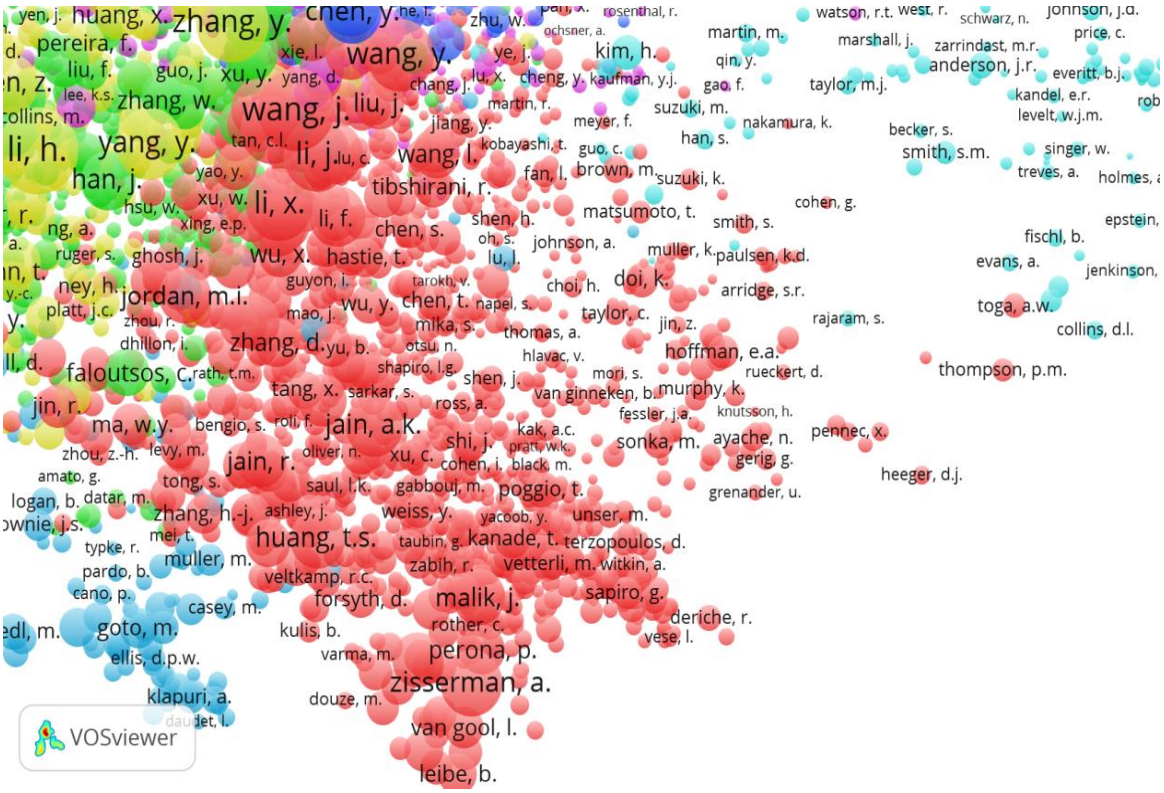

Image 2. Mapping author co-citation of cluster one

Author co-citation in library and information science with other disciplines is depicted in Image 2. In some parts cluster nodes are mingled with each other and this combination shows more cooperation in this scientific field with other scientific fields; such as, computer science, electrical engineering, information retrieval, multimedia information retrieval and psychology sciences and shows the proximity of these fields to one another. Also uncovers authors who have most connection with only the authors of their own field and they are less cited by authors of other fields.

Cluster two, shows the authors of computer science. As it is clear in picture 1 yellow colored nodes which are related to IR, have covered the green colored nodes which are related to computer science. Field of IR, due to its essence, uses computer science to achieve its goals. It is noteworthy that extension and development of this field is depended on science technologies. The drastic dependency and connection between these two fields can be observed from drastic approximation between clusters of nodes.

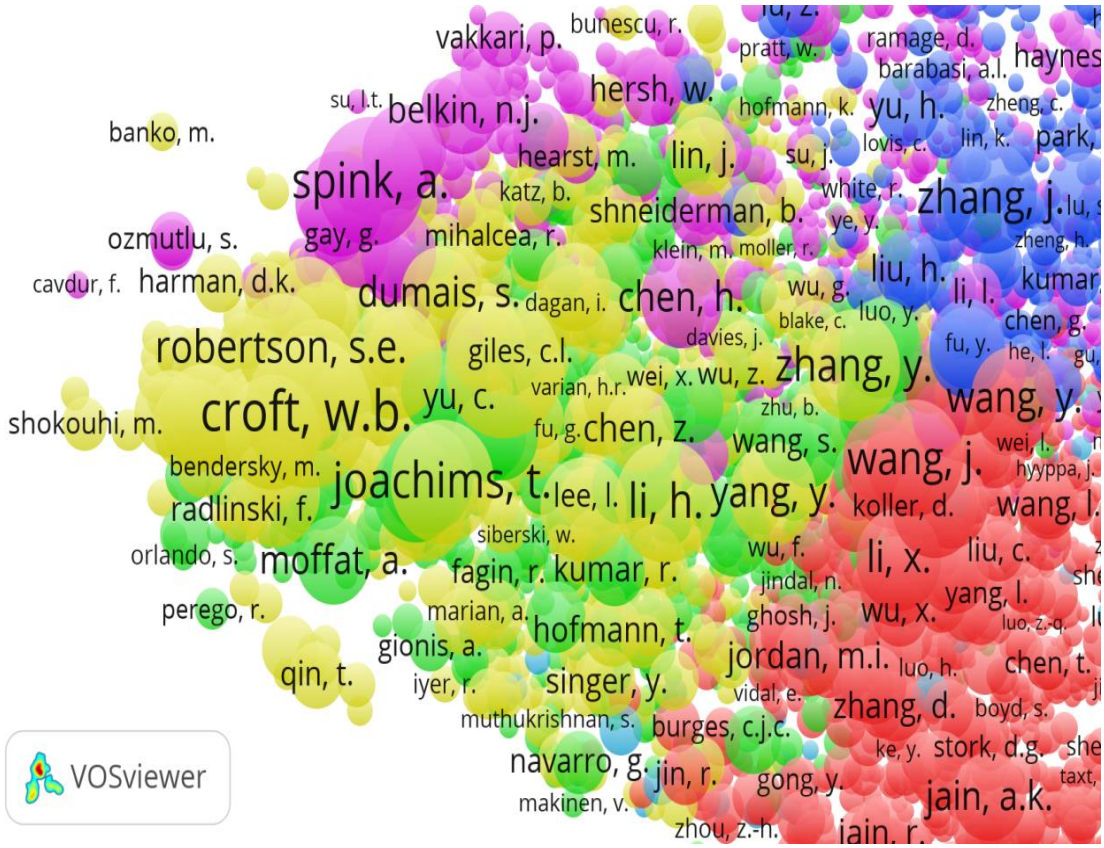

Image 3. Mapping author co-citation of cluster two and four

In Image 3, in addition to overlapping nodes of clusters two and four, proximity and intensity of relations between field of IR and other scientific fields can be observed. IR is also in close relation with fields of information behaviors (violet nodes), electronic engineering (blue nodes), software engineering, library and information science (red nodes) and multimedia information retrieval.

For more clarity of fields and relations between them, nodes are shown as cluster density visualization. In this model of visualization, density of each cluster is shown separately with specific color (Van Eck and Waltman, 2015). The biggest cluster of a map is shown by red color. The more density amount of a cluster decreases, the more intensity of color spectrum diminishes. In this visualization mapping density of each cluster can be observed. This mapping showed that field of psychology has the most distance with IR and computer and information behavior have the most relationships and proximity with cluster of IR.

Discussion

The Present study aimed to map, visualize, and determine subject trends in the field of IR through author co-citation network using SCOPUS (2005- 2018). Findings showed the top five authors in IR regarding citations are Wang, J. (2760 citations), Salton, G. (2643 citations), Zhang, Y. (2611 citations), Croft, W.B. (2450 citations) and Zhang, J. (2230 citations). A Research conducted by Ding et al. (1999), using Web of Science (1987-1997), introduced Salton, G., Belkin, N.J., Croft, W.B., Saracevic, T., and Robertson, S.E. as highly cited authors in IR. Accordingly, Salton, G. and Croft, W.B. have been recognized as two pioneers of IR. Regarding the most productive authors, Lin, H., Demner-Fushman, D.,

Cimino, J.J., Lu, Z. and Haynes, R.B. with 40, 37, 36, 36 and 33 articles, respectively are the five most productive authors of IR field in the years 2005-2018. In a similar study Rorissa & Yuan (2012), Thelwall, M., Spink, A., Nicholas, D., Jarveli, K. and Huntington, P. introduced as the five authors who had the most scientific productions in the years 2000-2009. Considering the reasons for the differences in the aforementioned results, one can point out the increasing growth of scientific productions and the increase of the activity of researchers in the field as well as differences in the type of search and citation databases being used. The comparison between highly productive authors and highly cited authors shows that authors with the most number of citations do not necessarily produce the most articles in IR field and this is in accordance with Daniali & Naghshineh (2014) research that all the highly productive authors were not among the highly cited authors.

Investigation of authors' affiliation showed that the majority are from the United States, China, the United Kingdom and Germany. So that 64% of the highly cited authors were from the United States and 26% of the 50 top authors were from China. Top-tier journals had been attracting more and more citations which are in favor of researchers. Results of the analysis of highly cited journals in IR showed that Nucleic Acids Research, IEEE Transactions on Pattern Analysis and Machine Intelligence, Journal of the American Society for Information Science and Technology (JASIST), Bioinformatics and Information Processing and Management, by receiving 12596, 8570, 8003, 7423 and 6354 citations, respectively are five highly cited journals. This finding is in accordance with the results of Ding et al. (2000). Also, Rorissa & Yuan (2012) introduced five highly cited journals of this field such as, Information Processing and Management, Communications of the ACM, Journal of the American Society for Information Science and Technology (JASIST) and the Journal of Documentation. On the other hand, analysis of the findings showed that Nucleic Acids Research, BMC Bioinformatics, IEEE Transactions on Image Processing, AMIA Annual Symposium Proceedings and the IEEE Transactions on Pattern Analysis and Machine Intelligence are five highly productive journals in the IR (2005- 2018). Comparison between the ten highly productive and highly cited journals showed that the Nucleic Acids Research, IEEE Transactions on Pattern Analysis and Machine Intelligence, In addition to producing the most articles in the IR field in the years 2005-2018, the most cited articles belong to these journals. It shows that IR journals are both effective and efficient a proxy for quality.

The findings also uncovered that the articles received the most citations from articles in the field of IR were published in highly productive and highly cited journals in this area. In addition to articles, we found four sources in the form of books that were categorized as frequently cited reference thanks to IR highly cited authors that chose them for publishing their findings.

Comparison of our results regarding highly cited publications with Rorissa &Yuan (2012) shows the differences and similarities. The works of Blei, Ng & Jordan (2003), Porter (1980), Baeza-yates, Ribeiro-neto, (1999), Salton, Mcgill (1983), Manning, Raghavan, Schutze (2008) are known as high cited works of the field of IR. But, they are ranked in different places based on Rorissa &Yuan's and our results. The growth process of IR productions fluctuates according to the publication years so that the most productions are in 2006 and the least in 2011.

Regarding institutions' activities in IR during 2005-2018, the results show Chinese Academy of Sciences, IEEE (Institute of Electrical and Electronics Engineers), CNRS (Centre

National de la Recherche Scientifique), Tsinghua University and University of Toronto were recognized as most active institutes among others. Rorissa & Yuan (2012) in their study depicted that University of Illinois at Urbana-Champaign, University of Maryland and Pennsylvania State University were highly productive institutes in IR.

Our analysis on the use of keywords considering IR showed that the most frequently used keywords are information retrieval (28906), human (19142), article (15114), information storage and retrieval (8102) and algorithms (5520). At least half of the keywords are related to computer science and the Internet, which is one reason for the impact of the field of computer on the advancement of IR. The findings of Beheshti & Tabatabaei, (2007) and Rorissa & Yuan (2012) confirm this result. It is noteworthy that three subcategories of computer science i.e. theories and methods, artificial intelligence, and software engineering are interwoven with IR as well as library and information science (Tabatabaei & Beheshti, 2007). It is the intrinsic interdisciplinary of IR. We found that the scientific relationship between IR and computer, medicine, engineering, biochemistry, genetic and molecular biology and social science is at top ranked. This is not in consistent with Rorissa & Yuan (2012) that showed relationship with computer, library and information science, engineering, telecommunications, and management are important for IR researchers.

Finally, author co-citation network in IR identified ten clusters. Analysis of subject trends of top authors in clusters showed that these clusters are assigned to the fields of library and information science, computer science, electrical engineering, information retrieval, information-seeking behavior, psychology, multimedia information retrieval, software engineering, ophthalmology and surgery.

The proximity and degree of clustering show intensity of the effect and the relationship between clusters. Therefore, clusters related to computer science, information-seeking behavior, electronic engineering, software engineering, information science and knowledge, and multimedia information retrieval have the most relationship with the field of IR. The authors co-citation mapping showed that the nodes in the field of IR and computer science are closer to each other, so that the yellow nodes (associated with the IR area), covered the green nodes (associated with the computer science). Analysis of clusters showed that psychology, ophthalmology and surgery are far from the area of IR. An important finding is that there is no linear correlation between number of citations and author co-citation. In this regard, we conclude that the more citation received, did not bring about co-citation within and outside clusters.

Conclusion

The study investigated IR and focused on its relationship with academic disciplines using ACA. Visualization of author co-citation map in IR determined ten clusters. Authors in these clusters were from different subject fields which were identified based on subject trend of top authors of each cluster from amount of co-citation outlook. This study depicted the intellectual structure of IR and showed ACA is a powerful tool for representing a scientific domain. Regarding knowledge management the ACA is very useful to represent implicit knowledge (Zavaraqi, 2010). Consequently, if one decides to quantify the relationship between co-cited authors as well as to map the structure of an academic discipline, ACA is suggested (Wang, Bu & Huang, 2018). Mapping knowledge domains envision the trends in academic disciplines. The respective information can be used to analyze the structure of

thematic, geographic, and institutional domain (de Moya-Anegon, Vargas-Quesada, Chinchilla-Rodriguez, Corera-Alvarez, Herrero-Solana, & Munoz-Fernández, 2005) in IR. Scientific domain development can be traced using ACA. Analysis of subject trends of top authors in clusters showed that these clusters are assigned to the fields of information science and Knowledge, computer science, electrical engineering, information retrieval, information-seeking behavior, psychology, multimedia information retrieval, software engineering, ophthalmology and surgery. We can conclude that Interdisciplinary nature of IR is a suitable context to determine the degree of academic domain interdisciplinarity. In this regard, providing co-authorship and interdisciplinary collaboration trends may help researchers, funding agencies, and policy makers in their decisions on intellectual structures in a given area.

Endnote

1. The information contained in this article was extracted from a master's thesis by the author, Fereshte Ehsanifar, entitled "Mapping, Visualizing and Determining Subject Trends in the Field of Information Retrieval by Using Author Co-Citation Network in SCOPUS (2005-2014)" submitted in partial fulfillment of the requirements for the degree of master of scientometrics at Regional Information Center for Science and Technology (RICeST)

Reference

Andrews, J. E. (2003). An author co-citation analysis of medical informatics. *Medical Library Association*, 91(1), 47-56.

Bayer, A.E., Smart, J. C. & McLaughlin, G. W. (1990). Mapping intellectual structure of a scientific subfield through author co-citations. *Journal of the American Society for Information Science*, 41(6), 444-452.

Berry, M. W., Dumais, S. T. & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM review*, 37(4), 573-595.

Borner, K., Chen, C. M. & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37(1), 179–255.

Chen, L. C., & Lien, Y. H. (2011). Using author co-citation analysis to examine the intellectual structure of e-learning: A MIS perspective. *Scientometrics*, 89(3), 867–886. Retrieved from: http://doi.org/10.1007/s11192-011-0458-y.

Cleverdon, c. w., Milis, J. and Keen, E. M. (1966). *Factors determining the performance of indexing systems* (2 Vols), Aslib Cranfield Research Project, College of Aeronautics, Cranfield

Daniali, S., & Naghshineh, N. (2014). Co-citation map of outstanding authors in the field of image retrieval. *Caspian Journal of scientometrics*, 1(2), 66-73. [in Persian]

de Moya-Anegon, F., Vargas-Quesada, B., Chinchilla-Rodriguez, Z., Corera-Alvarez, E., Herrero-Solana, V., & Munoz-Fernández, F. J. (2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Management*, 41(6), 1520-1533.

Dinet, J. (2014). Information Retrieval: Psycho-Ergonomic Approach. *Information Retrieval in Digital Environments*, 49-121. John Wiley & Sons, Inc.

Ding, Y., Chowdhury, G., & Foo, S. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *The International Information & Library Review*, 30(4), 367-376.

Ding, Y., Chowdhury, G., & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997. *Journal of Information Science*, 25(1), 67–78.

Ding, Y., Chowdhury, G. G., & Foo, S. (2000). Journal as markers of intellectual space: Journal Co-citation analysis of information retrieval area, 1987-1997. *Scientometrics*, 47(1), 55–73.

Ding, Y., Chowdhury, G., & Foo, S. (2001). Bibliometric cartography of information retrival research by using co-words analysis. *Information Processing and Managemnt*, 817–842.

Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, 57(1), 27–57.

Ingwersen, P. (1992). *Information Retrieval Interaction*. London: Taylor Graham.

Järvelin, K., & Vakkari, P. (1992). The Evolution of Library and Information Science 1965-85: A Content Analysis of Journal Articles. In: Vakkari, P. & Cronin, B. (Eds), *Conceptions of Library and Information Science: Historical, empirical and theoretical perspectives. London*, UK: Taylor Graham, 109-125.

Jeong, Y. K., Song, M., & Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics*, 8(1), 197-211.

Joachims, T. (2002). Optimizing search engines using click through data. *In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 133-142). ACM.

Kelly, D., & Sugimoto, C. R. (2013). A systematic review of interactive information retrieval evaluation studies, 1967–2006. *Journal of the American Society for Information Science and Technology*, 64(4), 745-770.

Ma, R., Dai, Q., Ni, C., & Li, X. (2009). An author co-citation analysis of information science in China with Chinese Google Scholar search engine, 2004–2006. *Scientometrics*, 81(1), 33-46.

Marshakova, I. V. (1973). A system of document connection based on references. *Scientific and Technical Information Serial of VINITI*, 6(2), 3-8.

McCain, K. (1990). Mapping author intellectual space: A technical overview. *Journal o the American Society for Information Science*, 41(66), 433-443.

Mooers, C. N. (1951). Zatocoding applied to mechanical organization of knowledge. *American Documentation*, 2(1), 20–32.

Nerur, S. P., Rasheed, A. A. & Natarajan, V. (2008). The intellectual structure of the strategic management, field: an author co-citation analysis. *Strategic Management Journal*, 29(3), 319-336.

Osareh, F., & McCain, K. (2008). The Structure of Iranian Chemistry Research, 1990–2006: An Author Cocitation Analysis. *Journal of the American Society for Information Science and technology*, 13(59), 2146–2155.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3):130-137.

Prime, C., Bassecoulard, E. & Zitt, M. (2002). Co-citations and co-sitations: A cautionary view on an analogy. *Scientometrics*, 54(2), 291-308.

Robertson, S. E. (1977). The probabilistic ranking principle in IR. *Journal of Documentation*, 33(4):294-304.

Rocchio, J.J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.) *the SMART Retrieval System: Experiments in Automatic Document Processing*, 313-323. Prentice Hall.

Rorissa, A., & Yuan, X. (2012). Visualizing and mapping the intellectual structure of information retrieval. *Information Processing and Management*, 48(1), 120–135. http://doi.org/10.1016/j.ipm.2011.03.004.

Salemi, N., Koosha, K. (2013). Co-citation Analysis and Co-word Analysis in Bibliometrics Mapping: A Methodological Evaluation. *Iranian Journal of Information Processing and Management*, 29(1), 253- 266.

Salton, G. (1981). The Smart environment for retrieval system evaluation–advantages and problem areas. In K. Sparck Jones,(Ed.), *Information Retrieval Experiment*, chapter 15, 316–329. Butterworths.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24 (4), 265-269.

Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799-813.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying Bibliography 215 and graphing specialties. *Science Studies*, 4(1), 17-40.

Subramani, M., Nerur, S. P., & Mahapatra, R. (2003). Examining the intellectual structure of knowledge management, 1990-2002–an author co-citation analysis. Management Information Systems Research Center, Carlson School of Management, University of Minnesota, 23.

Tabatabaei, N., & Beheshti, J. (2007). Delineation of Information Retrieval Research Area Using Input-Output Model. In Proceedings of the Annual Conference of CAIS/Actes du congrès annuel de l'ACSI. C. Arsenault and K. Dalkir, (Eds.) McGill University, Montreal, Quebec: 2007, pp. 1-7.

Taherian, A., & Osareh, F. (2011). Analysis and Mapping of Scientific Map of artificial intelligence in 1988 and 2008 in the Web of Science. *Journal of studies in library and information science*, 3(2), 31-68.[in Persian]

Teixeira, A. A. C. (2011). Mapping the (in) visible college(s) in the field of entrepreneurship. *Scientometrics*, 89(1), 1–36.

Van Eck, N. J., & Waltman, L. (2015). *VOSviewer Manual*. Retrieved from http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.3.pdf

Wang, B., Bu, Y., & Huang, W. B. (2018). Document-and Keyword-based Author Co-citation Analysis. *Data and Information Management*, *2*(2), 70-82.

White, H. D., & Griffith, B. C. (1981). Author cocitation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163-171.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-355.

Williams, R. V. (2010). Hans Peter Luhn and Herbert M. Ohlman: Their roles in the origins of keyword-in-context/permutation automatic indexing. Journal of the American Society for Information Science and Technology, 61(4), 835-849.

Zavaraqi, R. (2010). Author Co-Citation Analysis (ACA): a powerful tool for representing implicit knowledge of scholar knowledge workers. *Sixth International Conference on Webometrics, Informetrics and Scientometrics & Eleventh COLLNET Meeting*, Mysore (India), 19–22 October 2010.

Zhao, D. & Logan, E. (2002). Citation analysis using scientific publications on the Web as data source: A case study in the XML research area. *Scientometrics*, 54(3), 449-472.

Zhao, D. & Strotmann, A. (2011a). Counting first, last, or all authors in citation analysis: Collaborative Stem Cell Research Field. *Journal of the American Society for Information Science and Technology*, 62 (4), 654–676.

Zhao, D., & Strotmann, A. (2011b). Intellectual structure of stem cell research: a comprehensive author co-citation analysis of a highly collaborative and multidisciplinary field. *Scientometrics*, 87, 115-131.