Fall 9-29-2019

Mapping Intellectual Structure of Published Articles in Information Retrieval during 1983-2017.

Mozhdeh Salajeghe Prof.
Shahid Bahonar University of Kerman, Kerman, Iran, msala@uk.ac.ir

Adel Soleimani pro.
Shahid Bahonar University of Kerman, Kerman, Iran, adels2004@yahoo.com

Mahdieh Khazaneha MRS
Medical Science University of Kerman, Kerman, Iran, khazanehm@yahoo.com

Mapping Intellectual Structure of Published Articles in Information Retrieval during 1983-2017.

1. Adel Soleimani Nejad,
Affiliation: Shahid Bahonar University of Kerman, Kerman, Iran
Email: adels2004@yahoo.com

2. M Khazaneha,
Affiliation: Shahid Bahonar University of Kerman, Iran.
Email: khazanehm@yahoo.com

3. M. Salajegheh, responsible author
Knowledge and Information Science, Shahid Bahonar University of Kerman
Affiliation: Shahid Bahonar University of Kerman, Kerman, Iran.
Email: msalajgh@gmail.com
Phone: 09133439204

Abstract

Today information and communication cause the daily growth of published information. Studying all scientific production content and structures for specialist in different fields and publications is impossible. This study aims to analyze the articles regarding information retrieval based on the concepts of co-occurrence network analysis and centrality indicators published in Clarivate Analytics Web of Science[1] during 1983-2017. This is a descriptive study, using Scientometric approach. Its statistical population contains all articles related to Information retrieval in Clarivate Analytics Web of Science during1983-2017. The scientific research on Information retrieval starts in 2002. Based on the scientific map of countries, America, England, Canada and Singapore have the most articles in information retrieval field. Iran and Brazil have also been active in research on this field from 2012. The top authors of Articles in IR field articles during 1989-2017 are: Spink, Boregman, Chowdhury and Meado. In the analysis of IR field articles based on co-word anlaysis, 8 subject cluster were observed. Among them related to internal and external factors in information retrieval.

Key word: Scientometric method, Information retrieval, Search engine, Mapping intellectual structure

1. Introduction

Retrieval of information from the web is a subset of general information retrieval methods. Contextual information retrieval for searching information on the web is not a new idea but has distinct challenges when compared to either general information retrieval or non-

contextual information retrieval from the web. The research and development of information retrieval systems were growing along with the emerging of electro-mechanical searching devices. Along with the development of information and communication technologies, information retrieval researchers from the 1950s to the present day has done many studies, focusing on information retrieval field. The most obvious recent example of information retrieval field change is the rapid growth of mobile devices and social networks which help communities of users exchange their informal and unpublished information. Social networking has variety of areas and new tools for managing personal and social information (Dumais et al., 2003)

Information Retrieval (IR) is a broad term from subject scope point of view and is an interdisciplinary area of information science and computer science. It is related to monitoring, storage, processing and management, and searching and finding information in information systems. Recently, it is concerned with Internet/Web information retrieval or Web searching. The Web information retrieval, for indexing the fulltext of documents or a part of document uses the search engines and for classifying subjective web documents uses web directories (Baeza-Yates & Ribeiro-Netoz, 1999). In general, researches on information retrieval have been based on the status of retrieving requested information from databases and systems, particularly documents, images in web, etc.(D. Hong, Park, Lee, Shin, & Woo, 2005) .

Studies in information retrieval in different formats are constantly increasing because of information and communication technologies. Analyzing all scientific production contents and structures for specialist in different fields and publications of information retrieval is impossible. Today different scientometrics softwares help us to analyze different publication contents and structures. With analyzing research contents in information retrieval field the subjects which have been studied in this field and also not studied, will be extracted and analyzed.

Before emerging scientomeric and data mining programs content analysis of articles have been done by examining articles content. While, traditionally, the basis of IR researchers has been on creating systems which cover a variety of uses and needs. Now, it is necessary to focus more on the human, social, and economic impact of these systems as it has been done on the underlying algorithms and systems (Culpepper, Fernando Diaz, & Smucker, 2018)

Experts in a field counted the number of articles in a subject domain during a specific period of time. This method could be appropriate before 1970, because there was no information explosion like today, and also in that time scientometric tools was not developed like to day. Today, scientometric approach is used for discovering scientific contents and structure of scientific resources and mapping subject domain, co-citation, co-words and knowledge domain visualization. Knowledge domain visualization show the growth and structure of Scientific subject domain in publications like books, article, patents or grants. In the present research it is decided to analyses content structures of information retrieval field in web of science during 1983-2017 by scientometric analysis approach.

For analyzing the structure and development of research topics, co-word analysis methods will be applied based on co-word table, that includes factor analysis, cluster analysis, multivariate analysis and social network analysis. These methods help researchers to review subject field in a glance. It has an significant effect in understanding the value of academic discipline (Zong, etal. And Musgrove etal.).

Studying the evolution of subject fields of studies can help planners and managers, because they may use these frameworks in understanding the pattern of their services development.

Also, these evidences may provide a useful guideline for services designers to design characteristics of old and new services (Munan, 2018).
Awareness of subject growth in scientific areas is necessary for understanding the developments of topic domain in different sciences. Understanding subject domain of information retrieval studies which have been done during 1983 to 2017, help the researchers to know about different studies content and structure in this subject domain.

1.1. Research Questions

1. Based on co- word analysis, which countries have been working more on information retrieval field during 1983-2017?

2 .Based on the co-author analysis, who are the top authors in information retrieval during 1983-2017.

3 .Based on the co-word network analysis, the extracted clusters belong to which topics on information retrieval during 1983-2017?

4. what are the authors' subject category based on the centrality measures of co-author network analysis in information retrieval articles in Clarivate Analytics Web of Science during 1983 to 2017?

5.what are the authors' subject identifiers (ID) based on the Centrality measures of co-author network analysis in information retrieval articles in Clarivate Analytics Web of Science during 1983 to 2017?

6. What are the author's centrality indices of information retrieval in published studies in Clarivate Analytics Web of Science from 1983 to 2018?

Review of literature

Many researches has been done in other fields by scientometric approach like: library and information science (Liu, Hu &Wang, 2012); recommendation system in China (Hu, Zhang, 2015); robot technology in Korea (Lee, B. & Jeong, 2008); Intellectual structure of knowledge (Khasseh, etal., 2017). intellectual structure of information retrieval, (Rorissa & Yuan, 2012) ; research trends(Chen, et al , 2016). reverse logistics, closed loop supply chain management (kazemi, Modak & Govindan, 2018) and self-powered technology (Munan, 2018).
Application of hierarchical clustering led to the formation of 11 clusters representing the intellectual structure of iMetrics, including "Scientometric Databases and Indicators," "Citation Analysis," "Sociology of Science," "Issues Related to Rankings of Universities, Journals, etc.," "Information Visualization and Retrieval," "Mapping Intellectual Structure of Science," "Webometrics," "Industry–University–Government Relations," "Technometrics (Innovation and Patents), "Scientific Collaboration in Universities", and "Basics of Network Analysis" (Khasseh, et al., 2017). The growth of a subject domain usually, adheres of a model from adjusting state to mature state and Sometimes grow in readjusting state in the middle of evolving process(Chen, Tsutsui, Ding, & Ma, 2017) . Evolution models of subject domain are developed for discovering a subject and content of its publication. Although scientists studied scientific topic evolution for extending its fields and related subjects (Amoualian, Clausel, Eric, & Amini, 2016). Scientometrics and data mining approaches have explored the growth of scientific subject domains (Börner , Chen, & Boyack, 2005). Information retrieval takes some concepts and

ideas from computer science, Communication, Education, Engineering management, Business, Physic, Chemistry, Psychology and Nero science and give some concepts to it (Ding , Chowdhury, & Foo, 2000). The authors who have the most studies on information retrieval are the author who have the most co-operation with the other authors. Also they found that sub-fields of information retrieval include some new topics like, web information retrieval and user studies and topic development in comparison to 2000 subject areas (Rorissa & Yuan, 2012).

2. Research Methodology

Researcher method is content analysis with scientometric approach. Information retrieval topic is the target domain of this research for analyzing article contents in this topic. 1534 articles which were indexed in web of science during 1983-2017 are selected for content analysis. The search based is information retrieval related terms. article document type is selected for analyzing. Topics were extracted from articles titles and abstracts. Totally search, retrieval and processing of studied vocabularies can be summarized in 6 phases.

1. Collecting articles which were indexed in WOS in IR by the following search strategy:

   TITLE:("information retrieval model") *OR*TITLE: ("search engine") *OR* TITLE:("information retrieval") Refined by: WEB OF SCIENCE CATEGORIES: (INFORMATION SCIENCE LIBRARY SCIENCE) Timespan: 1983-2017. Indexes: SCI-EXPANDED, SSCI)
2. For collecting articles, Books review and Conference articles were removed and the remained articles have been selected as our sample. After searching and storing, retrieved articles have been extracted in 500 classes in text format and they have been sent to Excel.
3. Integrating vocabularies in Pre Map Raver software and excel; after indexing vocabularies with respect to the large amount of retrieved and natural language problems, it is needed to unify and somehow control vocabularies. So, need to a process was felt which can solve look of consistency in vocabularies and limit these large amount of vocabularies for unifying, based on information retrieval experts and sometime for identifying some concepts and full formats of abbreviation Google search engine was used.
4. Drawing conceptual map of authors and countries, co-operation and related subject area has been done by identifying the most used vocabularies and concepts. Mapping scientific structure was done based on, centrality indicators of information theories subject area and by Net draw, Ucinet, Excel, Vosviewer.
5. Authors group Co-operation rate of articles has been studied and articles based on the number of their authors were ranked.

3. Results

After retrieving 1534 record on information retrieval with scientometric programs and integrating data according to research objectives, data were analyzed by Co-authors and Co-occurrence of knowledge structure related to information retrieval field.

1. Based on co-word analysis, which countries have more articles on information retrieval in Clarivate Analytics Web of Science field during 1983-2017

Data of co-citation map of countries which are observed in figure 1 show that 48 countries are classified in 13 clusters. In this figure it is observed that, countries like, America, England, Germany, Spain and Canada have the most articles in information Retrieval. The most of the articles have been Published during 2002-2012 years. America in centrality of countries clusters have the most articles. In year 2002, America, England, Canada and Singapore had the most articles in IR. Also Iran, Brazil in 2012 have the most articles in information retrieval.

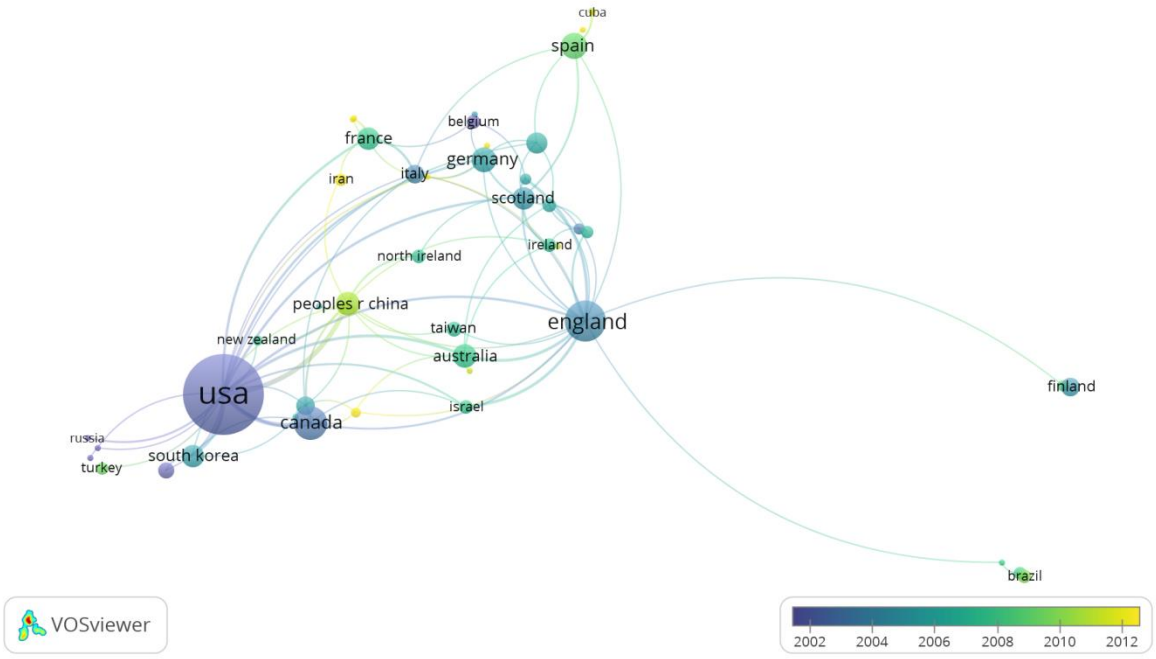

Figure 1. Collaborative countries in information retrieval during 1983-2017

No.	Authors	Frequency
1	Spink,a	**17**
2	Willet,p	**13**
3	Anonymos,	**12**
4	Cole,c	**11**
5	Jarvelinka	**11**
6	Boregman,CL	**10**

7	Ellis,D	**10**
8	Nottes,GR	**10**
9	Ding,Y	**9**
10	Losse,RM	**9**

Table1: Frequency of article published with authors in information retrieval

2. Based on the co-author network analysis, who are the top authors in information retrieval in Clarivate Analytics Web of Science during 1983-2017?

Existed links among authors is the cause of subject relationship and sharing among them. The most communication among authors is the reason for their most subject sharing and relationship and also the used colors are near to each other. The bigger node in the drawn map in mapped network is a cause of the significance of author in the created subject. Clusters in map show the most current and important authors who have important role in information retrieval subject in information science. In figure 2 it is observed that Spink, Boregman, Chowdhury and Meado are the centrality of this cluster. Based on article frequency in table 1, Spink, Willet, Anonymos, Cole and Jarvelinka have the most articles. According to social network analysis, Spink, Boregman, Chowdhury and Meado have the most contribution in Producing information retrieval in this network (figure 2).

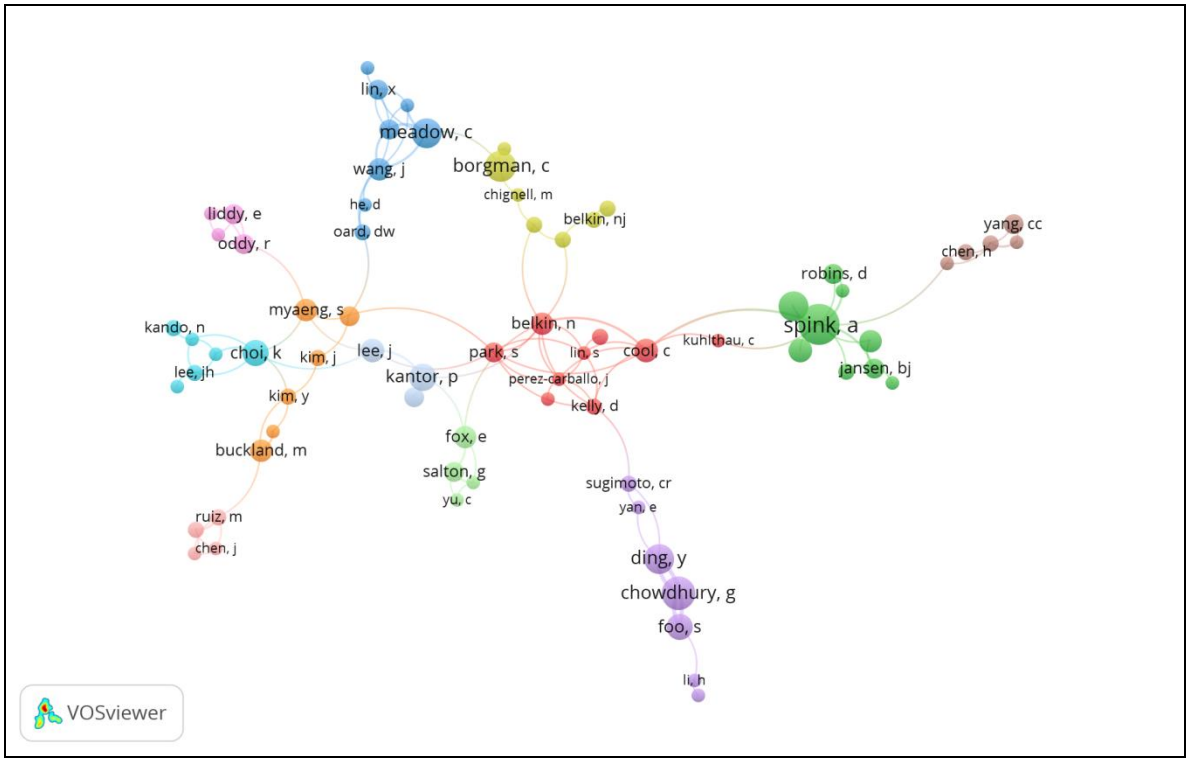

Figure 2: collaborative authors in co-authorship network analysis of information retrieval

3. Based on the co-word network analysis, the extracted clusters belong to which topics in information retrieval in Clarivate Analytics Web of Science during 1983-2017?

Subject studying in vocabularies show that vocabularies in information science, computer science, Health care services science and Medical information have the most indicators degree, have been studies centrality in IR in information science and show that IR studies with respect to indicator degree have the most application in library and information science, computer science, Health care services science and Medical information (figure 3 and table 2)

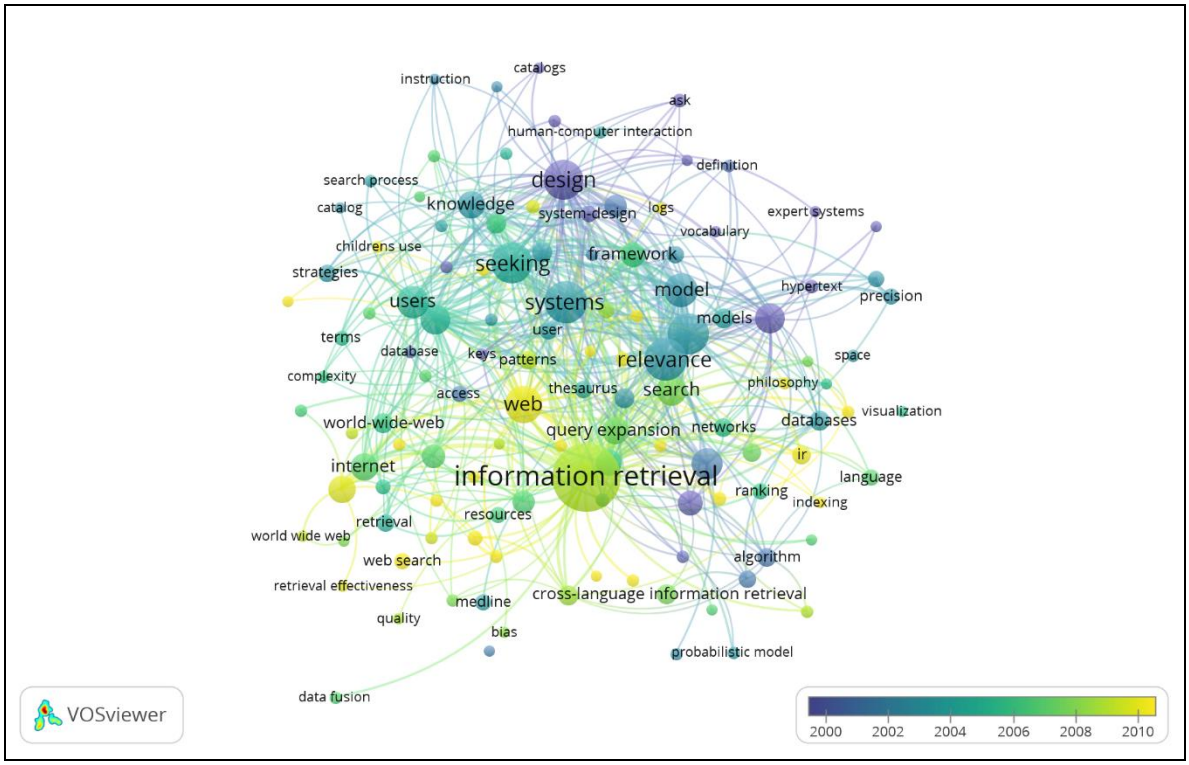

Figure 3. The co-word network analysis of information retrieval based on years.

Row	SC	Degree	Closeness	Eigenvector	Between
1	Information science & library science	1241	17	1	71.5
2	Computer science	1211	23	0.829	9.5

3	Health care sciences & services	105	28	0.027	0
4	Medical informatics	105	28	0.027	0
5	Geography	24	28	0.006	0
6	Physical geography	24	28	0.006	0
7	Business & economics	19	29	0.007	0
8	Education & educational research	15	30	0.006	0
9	Social issues	8	29	0.003	0
10	Social sciences - other topics	6	29	0.002	0
11	Communication	3	29	0.001	0
12	Government & law	3	30	0.001	0
13	Telecommunications	2	29	0	0
14	Arts & humanities - other topics	1	30	0	0
15	Science & technology - other topics	1	30	0	0

Table 2. Centrality indices of Subject Category on information retrieval in Clarivate Analytics Web of Science during 1983-2017.

4. what are the authors' subject category based on the centrality measures of co-author network analysis in information retrieval articles in Clarivate Analytics Web of Science during 1983 to 2017?

Regarding the centrality indices that are related to subject category, research results show that the most important categories in studies of information retrieval field conclude: information science, Computer science, Health care sciences and services and Medical informatics and geography.

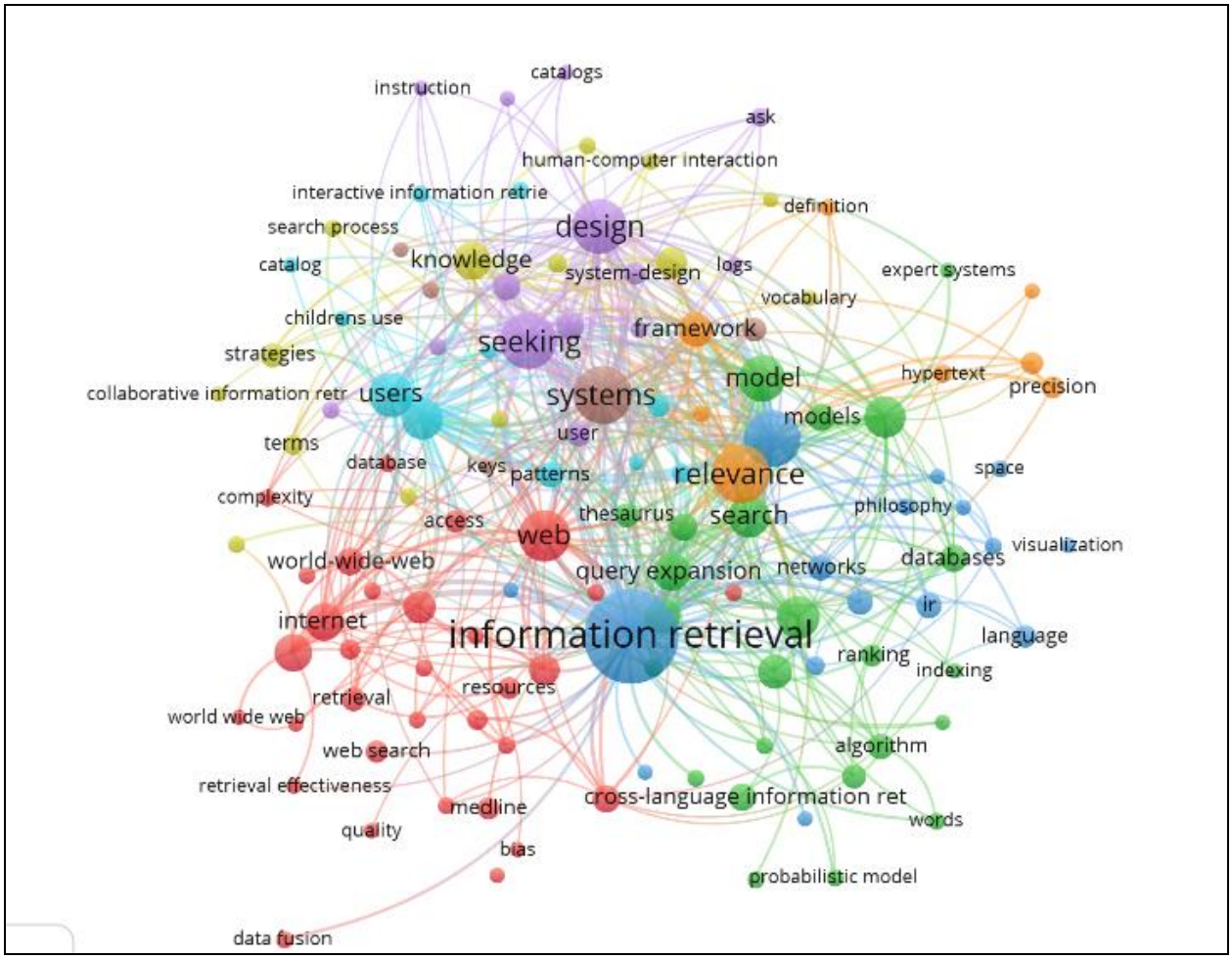

Figure4: The co-word network analysis of information retrieval

5.what are the authors, Identifier Description (ID) based on the Centrality measures of co-author network analysis in information retrieval articles in Clarivate Analytics Web of Science during 1983 to 2017?

Regarding the centrality indices, the most important identifiers in studies of information retrieval field are: Information science and library science, Computer science, information

systems, Computer science, interdisciplinary applications, Health care sciences and services and Medical informatics had the most effect on information retrieval. The results of this question have overlapped with results of question four. This shows that the researches were selected their research keywords form title and abstract (table 3).

Row	ID	Degree	Closeness	igenvector	etween
1	Information science & library science	1283	20	1	102.167
2	Computer science, information systems	1233	27	0.819	15.167
3	puter science, interdisciplinary applications	166	31	0.037	2.667
4	Health care sciences & services	140	33	0.027	0
5	Medical informatics	140	33	0.027	0
6	Geography	24	34	0.006	0
7	Geography, physical	24	34	0.006	0
8	Management	19	35	0.007	0
9	Education & educational research	15	36	0.006	0
10	Computer science, theory & methods	10	34	0.003	0
11	Social issues	8	35	0.003	0
12	Social sciences, interdisciplinary	6	35	0.002	0
13	Communication	3	35	0.001	0
14	Law	3	36	0.001	0
15	Computer science, artificial intelligence	2	35	0.001	0
16	Telecommunications	2	35	0	0
17	Humanities, multidisciplinary	1	36	0	0
18	Multidisciplinary sciences	1	36	0	0

Table3: Indicators of Centrality of authors in information retrieval articles in web of science from 1983-2017

6.What are the author's Centrality indices of information retrieval in published studies in Clarivate Analytics Web of Science from 1983 to 2018?

For studying performance and characteristic of existed authors nodes in social network in information science, indicators like centrality, Degree, betweenness, Closeness and eigenvector have been used in studying centrality indicators, authors like: Bruner, Hersh and Jarvelinka have the Highst degree indicator. this shows that these authors have the most co-writing with each other in betweenness indicator, Crestani, Sanderson, Hersh, Jarelin and

Clough have the highest betweenness. In closeness indicator authors like: Beheshti, Large, Criffon, Darmoni, Choi, Myaeng, Ding and Spink have the most closeness indicators in information retrieval. In Eigenvector indicator there is no author who has this indicator (table 4).

Row	Authors	Degree	Closeness	Eigenvector	Between
1	Ding,y	19	28091	0	888.5
2	Hersh,w	17	27597	0	2425
3	Jarvelin,k	17	27687	0	1952.333
4	Spink,a	17	28065	0	995.5
5	Darmoni,sj	16	28650	0	43.5
6	Belkin,nj	15	28032	0	1074.5
7	Choi,ks	15	28399	0	320
8	Crestani,f	14	27419	0	4325
9	Foo,s	14	28139	0	489.5
10	Griffon,n	14	28652	0	17.5
11	Myaeng,sh	14	28391	0	442
12	Sanderson,m	14	27487	0	3715.833
13	Clough,p	13	27420	0	1938.5
14	Hersh,wr	13	27829	0	600.4
15	Jones,cb	13	27454	0	411
16	Large,a	13	28683	0	20.25
17	Lee,jh	13	28339	0	288
18	Zhang,j	13	28338	0	314
19	Beheshti,j	12	28683	0	20.25
20	Berrios, dc	12	28720	0	11.5

Table4: Centrality indices of authors on information retrieval in Clarivate Analytics WOS

4. Discussion and conclusion

The scientific research on Information retrieval starts in 2002. Based on the scientific map of countries, Information retrieval America, England, Canada and Singapore had the most articles. Iran, Brazil have also been active in research on this filed from 2012. The top authors in IR field in Clarivate Analytics during 1989-2017 are that Spink, Boregman, Chowdhury and Meado. The analysis of IR also showed that based on co-word analysis there are 8 clusters. Figure (4).

According to data mining and scientometric analysis of information retrieval studies, as is observed in figure 1, the first cluster in the information retrieval studies is information retrieval (IR) in blue color. In this cluster world wide web is the topic which has the most articles. In studies which have been done in web in relation to IR cluster, subjects like: network, database, users' studies, retrieval effectiveness, users demand and queries, access information seeking, evaluating image search, and complexity are sub-fields of Web as a main cluster. The possible reason for appearing web as subfield of IR studies, is that today, web is a place where people produce, search and retrieve their information, of every kind, and in different format and web as a multimedia. So web has the main role in producing, Searching and retrieving IR articles. First of all, users need to hold a topic for searching in web, which we call search purpose. Sometimes short phrases and a few words submit the text to search engines. List of ranked URLs are returned with descriptions of these pages after users submitted their queries and these relevance pages show with queries how these are ranked by search engine. The system then ranking and merging the links we obtain from different search engines for the query we give.

One of relevant web is metasearch engines. The effectiveness of a metasearch engine is determined by the quality of the results and it returns and in response to user queries. The number of models of search engine being used is increasing, and so is the need for efficient mechanisms to search them. Different existing search engines could be used for this purpose, but they lack to properly search models, mainly they are strongly focused on text-based search mainly. For example one of model, Moogle, is used like metamodeling information to involve richer indexes and to allow more complex queries (Lucrédio, Fortes, & Whittle, 2012).

The second cluster of IR which according to scientometric content analysis, has been observed, is network, it can be because of internet network and web uses as a base for producing, searching and retrieving of information. Recently without network and web, information retrieving already is impossible. So, every study with information retrieval has web and Internet network subfield. One of the other subfield of IR which has seen in information retrieval cluster is database, databases as a collection of information resources which are stored, searched, and retrieved by DBMS, have the task of retrieving information. Therefore, it is natural that along with information retrieval cluster, database appear as a subfield. Ding et al (2000). A spatial database is defined as a pool of integrated and structured geospatial data, which is a model of reality, and from which data may be retrieved to provide useful information to users for instance GIS based on spatial database. Results show that today, information retrieval is internet- based and online.

IR Services is designed for meeting users' information needs. IR systems target is meeting users' demands and needs. Different information needs may lead to different behaviors. So everywhere IR is studied, users are a subfield of it. Users by information Retrieval systems

will find their responses to questions like where they can retrieve, how they can find my information and sources, and Which information or information sources are related.

The third cluster related to information Retrieval which is observed in figure (4) is databases in green color. The subfields of this concept are boolean operators, algorithm, language information, document retrieval, exports system, sampling, natural language, processing models, possibility models, query search Extending, ranking, relevance, system selecting, text retrieval thesaurus co-occurance distributed the activity of optimizing individual websites and webpages of information retrieval to get higher page rank and relevance page in the search results. Webpage content should be monitored keywords and through referrals. through proper keyword selection and link building should be taken into account when developers and designers work on a search engine or software that related to IR (Khan & Mahmood, 2018).

After web, databases are the places where users can find the organized information near to tradition information systems formats. Databases conclude of sources like books, article, dissertation, references sources, journals, multimedia sources and the other sources which can be found in traditional libraries. In most of databases, users should know something before retrieving information like search strategies, search engines search type and browsing search type, how to develop their queries for searching and selecting databases . Patel has shown that if the database permits quantum queries, then mere digitization is sufficient for efficient search for one desired item (Vijaya, Raju, & Ray, 2016). In addition, Data mining plans have been extremely utilized for deleting non-paltry data facts from such large volume of data. Overall, data mining in database is discovering many scientific facts in web site or database. Also most of search engines are web-based for Biomedical like: Biocarain, Crescendo, Sagace, Integrome DB and etc. that enable users to retrieve information from a range of biological databases.

Another item in relation to databases which is shown, is thesaurus, this item related to Q&A system. Quality and the accuracy or Q&A system can improve result of search. Classification of words is important when placed in graph layer. Therefore, the large-scale automatic structure is realized from the different thesaurus to each subject knowledge graph. Intensive research has been conducted in this area during the past decade. However, most researchers failed to notice that ignoring the semantic importance of certain feature terms might also contribute to low classification accuracy(Wang, Huang, Yang, & Li, 2012).

The forth cluster which is observed as a concept for information retrieval is information retrieval, which was shown in blue color. The terms related to information retrieval cluster are information retrieval systems, language, networks, creating philosophy, semantic web, visualizing, co-citation, analysis, bibliometric and documents. For retrieving information, it is necessary to have information retrieval systems, where facilities are provided for users' information searching and retrieving. Definition of natural language and controlled languages and kinds of language should be defined, the information retrieval should be based on world wide web network, so the users could retrieve their information needs in different formats like texts image and multimedia.

Most of soft wares of IR, advanced legal documents and knowledge management systems, are based on legislative XML and ontologies. Also the precision and recall metrics are used to evaluate the performance and the keyword-based search system. In recent years, the use of

visualization techniques increased and they have new challenges and broad discussions in the area of data storage and retrieval. Clustering is important subject of visualization and it has different kinds and the most important is hierarchical clustering algorithms, partition algorithms and fuzzy algorithms. These clusters are being used in retrieval information, thesaurus, Q&A systems, clustering search engine and Scientometric studies.

Users for seeking their information needs, need to a personal environment for operations like: searching, retrieving, making queries, storing their retrieved information and totally they need to this environment as a user interface for interacting to information retrieval systems. If users learn how to seek information needs formally, they can retrieve their information quickly and easily. After retrieving information users have a feedback, if the feedback meets their information needs, they stopped information seeking, and if their feedback doesn't meet their information needs, they should repeat seeking strategies and searching queries again. For example, investigating programming learners' information seeking behaviors in online, and provide visual navigational support to facilitate information seeking. This information is extracted by model sequence pattern through mining techniques (Lu & Hsiao, 2017).

In addition to traditional search engine, Citeseerx as a new search engine is explained. By this new search engine, users can access to documents full text, automatically extract metadata and citation context of papers. In Citeseerx metadata are provided through Open Archive Initiative (OAI) service interface and Amazon Charges based on usage. Citeseerx can extract and indexing different parts of an article like tables, figures which were not extracted in traditional search engines. Citeseerx can provide a framework for digital library search engine for staying on similar sites. Artificial Intelligence techniques are used in Citeseerx parts like; document classification, simulation, extracting metadata automatically, and author's disambiguation. Citeseerx is a model for interaction between information retrieval systems and users.

According to data mining and scientometric analysis on information retrieval studies content, fifth cluster which is in purple color, is information seeking. Information seeking is a broad term and its narrow terms are information seeking strategies, users' instruction, users' interface, system design, online catalogs, feedback of search, queries, user and password and orientation. These are information seeking behavior subfields because a user for seeking his/her information need to have a query for gaining to his/her information needs, after having a query they need to search strategies for searching and retrieving their information needs quickly and easily. Every user for securing communication with information retrieval systems, need user and password account and a personal account.

In interaction retrieval systems, there are different search strategies. Changes in information retrieval are approaches which consider human aspects in information retrieval. Cross-section of users and systems in interactive information retrieval which is resulted to a research specialty. The objective of this cluster is helping users to understand systems and meeting information needs through interacting with systems. In interactive information retrieval systems, resources ranking will study through citation analysis like Bradford. In interactive retrieval systems there are queries like asking question and HQE Models or Hybrid Queries which are extension model. This model can be used in different frameworks which are based on communication rules and its combination with external knowledge. It can be used for developing queries models. Most of systems are searched alone, but today modelling methods

based on content, resolved this problem, one of this models is PRE or Pseudo-Relevance Feedback framework. It is necessary to mention that in interactive retrieval systems based on graph like GIBIR which is one of the effective method for retrieving Bibliographic information, this system is suggested. Bibliographic information as a network shows and provides an asking question interface based on form. In this system, users can create their demands interactively as graph.

The sixth cluster for information retrieval studies is users' interaction information retrieval systems. User for interacting information in retrieval systems with other users, need user interface and web environment for share their thoughts, and information and interact with information retrieval systems for searching and retrieving their information needs. Interaction information retrieval need to communication and information technology like: web search engines, ontology, applications, knowing how to communicate, be familiar with information science and making and revising their queries. In information seeking behavior' searches there are vocabularies ambiguity. These ambiguity is solvable through connecting to thesaurus like Wordnet which is using in space detection retrieval model.

Seventh cluster in orange color for information retrieval studies which are analyzed by scientometric analysis is recall and precision. Users for finding relevant information and also recall need to know, searching and evaluating frameworks, making logical relationship between searched terms, they should know how to define concepts, and how to evaluate search results according to evaluation standards.

In this cluster, Hypertext (Hypertext Markup Language) is observed. HTML pages include noisy data such as advertisements, navigation menus and privacy notes. Moreover, some pages cover several different topics that making it difficult to identify the most relevant to the user. The approach combines page clustering techniques based on Document Object Model-based representations for acquiring evidence about relevant correlations between text contents. This evidence is exploited for better filtering out irrelevant information and facilitating the construction of interest profiles (Gasparetti, 2017).

The eighth cluster in mustard color, which is extracted from information retrieval studies is search process, which itself as a main topic has some subfields like user interface, libraries, information needs, personal differences, computer and human interaction, search strategies, terms, keywords, world wide web and online. Without search process, users cannot retrieve any information. For searching information, they should know how to search as a process. Libraries are systems where users can search resources by traditional ways like loan. They can also search information in digital libraries which now are the most usable libraries. In this research Information systems like digital libraries is not observed as a cluster or even as a subfield. Digital libraries were designed for users to interact with computer system to meet their needs. For searching some instruments like ontology, thesaurus, search strategies, online resources and information, they are needed, which are not seen in eighth cluster. An environment like world wide web which help users to be online and in contact with other people also is needed which is not observed in subfields of this cluster. According to personal differences, people have different needs. systems like information centers and libraries which aren't observed among systems' subfield in this research. Systems like libraries and information centers as information retrieval system, are crucial today, these systems for helping users to access their information should have necessary technologies and equipment

for following the demands of user. In Libraries and information centers, librarian and information professionals which are not observed in cluster eight are users' interface with systems.

4.1 Conclusion

Research findings show that the top researchers in information retrieval during 1983-2017 in WOS are Spink, Boregman, Chowdhury and Meado. America, England, Canada and Singapore countries have the most articles. Iran, Brazil have also been active in research on this filed from 2012. The analysis of IR also showed that based on co-word analysis there are 8 clusters (figure, 4). The eight clusters are showing the topics that have been studied in information retrieval as a broad subject more than the other topics. These topics include: world wide web, internet network, databases, user' information needs, information retrieval, search engines like engine citeseerx, Information seeking behavior, Interaction retrieval system, recall and precision, systems, search process. Fields in eight clusters show that all eight clusters are related to information retrieval, there is no any anonymous field which don't related to IR. Some fields are common in two or three clusters like queries in cluster on, five and six. Despite the significance of some topics like libraries and information centers, Indexing, digital libraries, information sources and links. these topics has been shown in narrow or related topics, Not as main topics in IR field. It may be because of missing them in the keywords and abstracts of studies which has been analyzed or the other reason which need to do more studies

References

Amoualian, H., Clausel, M., Eric, G., E. , & Amini, M. R. (2016). *Streaming-LDA: A Copula-based Approach to Modeling Topic Dependencies in Document Streams.* Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining San Francisco, CA, USA. DOI: http://dx.doi.org/10.1145/2939672.2939781

Baeza-Yates, R., & Ribeiro-Netoz. (1999). *Modern Information Retrieval,* : ACM Press. .

Börner , K., Chen, C., & Boyack, K. (2005). Visualizing knowledge domains. . *Annual Review of Information Science and Technology, 37*(1), 179-255.

Chen, B., Tsutsui, S., Ding, Y., & Ma, F. (2017). Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval. *Journal of Informetrics, 11*(4), 1175-1189. doi: https://doi.org/10.1016/j.joi.2017.10.003

Chen, X. Chen, J. Wu, D., Xie, Y. & Li, J., (2016). Mapping the research trends by co-word analysis based on keywords from funded project. Procedia Computer Science 91 (2016) 547 – 555. https://doi.org/10.1016/j.procs.2016.07.140

Culpepper, J. S., Fernando Diaz, F., & Smucker, M. D. (2018). Research Frontiers in Information Retrieval. Report from the Third Strategic Workshop on Information Retrieval in Lorne

ACM SIGIR Forum, 52(1), 37-90. doi: https://doi.org/10.1145/1113343.1113344. https://doi.org/10.1145/2215676.2215678

D. Hong, D., Park, Y., Lee, J., Shin, V., & Woo, W. (2005). *Personalized Information Retrieval Framework.* Paper presented at the the First Internaltional Workshop on Personalized Context Modeling and Management for UbiComp Applications.

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), 155-173. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1031>3.0.CO;2-B.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., & Robbins, D. (2003). *Stuff I've seen: a system for personal information retrieval and re-use', in* Paper presented at the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, Canada. .

Gasparetti, F. (2017). Modeling user interests from web browsing activities. *Data mining and knowledge discovery, 31*(2), 502-547. doi: 10.1007/s10618-016-0482-x.

Hu, C.P., Hu, J. M. Deng, S.L. & Liu,Y. A. (2013) co-word analysis of library and information science in China. Sciemtometrics, 97(2): 369–382. https://*doi*.org/*10.1007/s11192-013-1076-7*.

Hu, J. Zhang, Y. (2015).Research patterns and trends of Recommendation System in China using co-word analysis, Information Processing & Management, 51: 329-339. http://dx.doi.org/10.1016/j.ipm.2015.02.002

Khan, M., & Mahmood, A. (2018). A distinctive approach to obtain higher page rank through search engine optimization. *Sādhanā, 43*(3): 43. https://doi.org/10.1007/s12046-018-0812-3.

Kazemi, N., Modak, N. M. & Govindan, K. (2018). A review of reverse logistics and closed loop supply chain management studies published in IJPR: a bibliometric and content analysis. International Journal of Production Research. DOI: 1080/00207543.2018.1471244

Khasseh, A.K. Soheili, F. Sharif Moghaddam, H. and Mousavi Chelak, A.(2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. Information Processing & Management. 53 (3) 705-720. DOI:10.1016/j.ipm.2017.02.001

Lee, B. & Jeong, Y.I. (2008). Mapping Korea's national R&D domain of robot technology by using the co-word analysis, Scientometrics, 77 (2008) 3-19. DOI: 10.1007/s11192-007-1819-4.

Liu, G.Y. Hu, J.M. Wang, H.L. (2012). Co-word analysis of digital library field in China, scientometrics, 91 (2012) 203-217. https://doi.org/10.1007/s11192-011-0586-

Lu, Y., & Hsiao, I.-H. (2017). Personalized Information Seeking Assistant (PiSA): from programming information seeking to learning. *Information Retrieval Journal*, *20*(5), 433-455.

Lucrédio, D., Fortes, R. P. d. M., & Whittle, J. (2012). MOOGLE: a metamodel-based model search engine. *Software & Systems Modeling*, *11*(2), 183-208.

Munan, Li (2018). Visualizing the knowledge profile on self-powered technology. Nano Energy 51, 250-259. https://doi.org/10.1016/j.nanoen.2018.06.06.

Musgrove, P.B., Binns, R., Page-Kennedy, T., Thelwall, M. A method for identifying clusters in sets of inter-linking Web spaces, Scientometrics, 58 (2003) 657-672.

Rorissa, R., & Yuan, X. (2012). Visualizing and mapping the intellectual structure of information retrieval. *Information Processing & Management*, *48*(1), 120-135. doi: http://hdl.handle.net/1951/63232

Shrivastava, P., Ivanaj, S. and Ivanaj, V. (2016) 'Strategic technological innovation for sustainable development', Int. J. Technology Management, Vol. 70, No. 1, pp.76–107.

Vijaya, P., Raju, G., & Ray, S. K. (2016). Artificial neural network-based merging score for Meta search engine. *Journal of Central South University*, *23*(10), 2604-2615.

Wang, B.-k., Huang, Y.-f., Yang, W.-x., & Li, X. (2012). Short text classification based on strong feature thesaurus. *Journal of Zhejiang University SCIENCE C*, *13*(9), 649-659.

Wu, D.S., Xie, Y.J. Dai, Q.Z. & Li, J.P. (2016). A systematic overview of operations research/ management science research in Mainland China: Bibliometric analysis of the period 2001-2013, Asia-Pacific Journal of Operational Research, forthcoming.

Zong, Q.J., Shen, H.Z., Yuan, Q.J. , Hu, X.W., Hou, Z.P., Deng, S.G(2013). Doctoral dissertations of Library and Information Science in China: A co-word analysis, Scientometrics, 94 : 781-799.

.