Qualitative and Quantitative Methods in Libraries (QQML) 10,4 : 623-647, 2021

Global Visualization and Knowledge Mapping in the Field of Information Retrieval (IR): A Bibliometrics Analysis

Sk Sofik[1] and Ziaur Rahman[2]

[1]Research Scholar, Department of Library and Information Science, Rabindra Bharati University, 56A, BT Road, Kolkata, -700050. West Bengal, India E-mail- sofikrbu@gmail.com

[2]Assistant Professor, Department of Library and Information Science, Rabindra Bharati University, 56A, BT Road, Kolkata, 700050. E-mail-- ziaur.rahman@rbu.ac.in ORCID ID – https://orcid.org/0000-0002-3393-8690700050.

Abstract Information retrieval (IR) is the procedure of obtaining information system resources relevant to an information need from collecting those resources. Information Retrieval (IR) has experienced tremendous growth over the years and has become the dominant form of Information access. The prime purpose of this study is to explore global scientific publications regarding Information Retrieval (IR) and mapping the recent global trends and visualization in this innovative research domain, thus provide future research directions based upon dynamic observations. With a bibliometric analytical framework, the study explores the Web of Science core collection database covering published materials from 2006 to 2020. The article also developed a graphical visualization and knowledge mapping of the bibliographic literature by using ScientoPY, Biblioshiny, Histcite, and VOS-viewer to map journals, keywords, and institutions with bibliographic co-occurrence of terms, topic, and co-citation analysis. The study carefully evaluates the current scenario of IR that will lead the researcher to form innovative initiatives to facilitate effective and further research.

Keywords: Bibliometrics, Scientometrics, Citation Analysis, Information Retrieval, Trend Analysis, Research Productivity

Received: 4.7.2021 / Accepted: 17.12.2021 ISSN 2241-1925

1. Introduction

Information retrieval (IR) forms a significant component of both the interdisciplinary field of information science and the discipline of computer science. Generally, IR is concerned with the representation, storage, organization, and access of information through information systems. (Rorissa & Yuan, 2012) IR is the process by which a collection of data is represented, stored, and searched for knowledge discovery as a response to a user request (query) .this process involves various stages that represent data and return relevant information to the user. The intermediate set includes filtering, searching, matching, and ranking operations. The main goal of an information retrieval system (IRS) is to "finding relevant information or a document that satisfies user information needs" (Roshdi & Roohparvar, 2015)

(Saracevic, 1999)considers the problems tackled by information retrieval practitioners at the core of information science, although information science is much more than IR. Although the early years of IR research focused on retrieval systems and methods, researchers have since incorporated the cognitive, interactive, and contextual aspects of information seeking and searching into IR research and system design. In doing so, the IR subfield expanded to address issues related to users,use, situations, context, and users' interactions with systems. Greater issue inclusion resulted in the splitting of the retrievalcluster into two subclusters: systems-centered and user-centered. (Saracevic, 1999)

Information retrieval is a long-established subfield of library and information science. Since its inception in the early- to mid -the 1950s, it has grown as a result, in part, of well-regarded retrieval system evaluation exercises/campaigns, the proliferation of Web search engines, and the expansion of digital libraries. Although researchers have examined the intellectual structure and nature of the general library and information science field, the same cannot be said about the subfield of information retrieval. (Rorissa & Yuan, 2012)

Information retrieval systems are divided into two categories: Database systems that store and retrieve structured data and Search Engine systems that store and retrieve textual documents. Today, information can be represented in many different formats: structured data, text, image, audio, video, RSS feeds, tags, people, relations, etc. For the last five years, IR systems have been developed to cope with these changes and differentiate between structured and unstructured data. However, more recently, the Semantic Web community started pushing hard for omnipresent data, thus no longer distinguish between structured and unstructured data.(Zhuhadar, Nasraoui, Wyatt, & Romero, 2009).

2. Literature Review

(Sheikhshoaei, 2020)explores the intellectual and cognitive structures of information retrieval (IR) in the medical sciences through science mapping. The results showed that the similarity index increased over time from 0.43 to 0.71. Analysis of the findings shows that Similarity measures, Expert systems, Concepts, Experience, Answers, and Multi-model IR clusters are considered

mature and completely centralized clusters in the first quarter of the strategic chart. The effectiveness of scientific documents on answering clinical questions and focusing on health professionals' information behaviors has increased compared to search methods and tools.(Safder & Hassan, 2019) proposed a novel deep learning-based feature engineering approach that improves search capabilities by mining algorithmic-specific metadata from full-text scientific publications. Typically, standard term frequency-inverse document frequency (TF-IDF)-based approaches function like a 'bag of words model and thus fail to capture either the text's semantics or the word sequence. In this work, we designed a semantically enriched synopsis of each full-text document by adding algorithmic-specific deep metadata text lines to enhance the search mechanism of algorithm search systems. These text lines are classified by our deployed deep learning-based bi-directional long short-term memory (LSTM) model. (Bar-Ilan, 2017) investigates issues by submitting the same query to different databases relevant to the query topic. Information was retrieved f from three databases: ACM Digital Library, WOS (with the Proceedings Citation Index), and Scopus. The ACM Digital Library data were retrieved from the more comprehensive ACM Guide to Computer Literature that includes non-ACM data and covers the major journals in information science. Altogether, 9050 items were retrieved, out of which 5591 (62%) items were retrieved by a single database only, and 1059 (12%) items were located in all three databases. There are great variations in the citation counts as well. (Bar-Ilan et al., 2016) the study analyzed that bibliometrics and information retrieval to discuss how each of these two important areas of information science can help inform the research of the other. There is a growing body of literature that capitalizes. (Glänzel,2015) examines the interrelation between information retrieval and scientometrics pointed at possible synergy effects provided by some recently developed bibliometric methods in the context of subject delineation and clustering. Examples of specific search strategies based on traditional retrieval techniques and bibliometric methods illustrate this approach. Special attention is paid to hybrid techniques and the use of 'core documents.' The latter is defined merely based on bibliometric similarities but have by definition properties that make 'core documents' interesting and attractive for information retrieval.(Rorissa & Yuan, 2012) investigate that the information retrieval intellectual landscape through visualizations of citation behaviors. Citation data for ten years (2000–2009) were retrieved from the Web of Science and analyzed using existing visualization techniques. Our results address information retrieval's co-authorship network, highly productive authors, highly cited journals and papers, author-assigned keywords, active institutions, and importing ideas from other disciplines.(Ding, Chowdhury, Foo, & Qian, 2000) examines that the results obtained from specific bibliographic research can be applied to a real search environment and enhance the utility of an information retrieval session for all levels of end-users. In this respect, a Web-based Bibliometric Information Retrieval System (BIRS) has been designed and created to assist the end-users in understanding their search domain, formulate and expand their search queries, and visualize the bibliographic research results.

A preliminary user evaluation study of BIRS revealed that users generally found it easy to form and expand their queries. BIRS helped them acquire useful background information about the search domain. (Ding, Foo, & Chowdhury, 1998) this study analyzed the collaborative pattern of the Information Retrieval (IR) research field using coauthored articles retrieved from *the Social Science Citation Index* for 11 years from 1987 to 1997. This study reveals a perceptible upward trend of collaborative IR research, with the results of these research efforts been reported in all major core IR journals. The inter-disciplinary and intra-disciplinary scholarly communications in collaborative research are evidence and cover broad areas like psychology and computer and medical sciences.(Quoniam, Balme, Rostaing, Giraud, & Dou, 1998)investigate that used Zipf s law to qualify all the keywords of documents in a data set. Used this qualification to build a graphical representation of the resulting indicator in each document. The graphical resolution leads to a document dispatch in a three-dimensional space. This graphical representation is used as an information retrieval tool without using any keyword. The presentation of a case study is internet available. The graph is drawn in Virtual Reality Markup Language (VRML), allowing a dynamic picture linked to a Database Management System (Free wais). The experimentation was drawn to get a first impression of documents data set by querying without any keyword. (Ying Ding, 1993) analyses that inter-journal citations permit objective evaluation of the network of journalsin and around a discipline. This study aims to map the journal's role in the scholarlycommunication process of the IR field and their changes over time by usingjournal co-citation analysis between 1987 and 1997. The results of this study show that (i) the scholarly communicationpositions of the journals which are from different areas play an important role in the IR field, (ii)the field of IR is a mature field because the journals used for research communication remainedquite stable during the study period. This research also identifies those LIS journals which arehighly used for research communication by IR researchers.

3. Objectives of the Study

The main objective of this study is to analyze the global research performance in the field of IR as reflected in the publication and citation output during 2006-2020. In particular, the study focuses on the following aspects

- To study the year-wise growth of publications and citations.
- To identify the document type and language-wise distribution
- To determine the most productive countries, organizations, and authors.
- To identify the preferred journals of researchers in IR.
- To identify the highly influential research papers concerning citation and average citation per year on IR.
- To explore the most frequently used author keywords and co-occurrence of author keywords network in IR.
- To explore the most frequently used all keywords in IR.

- To explore the trends topics,clustering, and scaling of key concepts in IR.
- To find out the collaboration network (Author, Institutions, and Country) of IR researchers.
- To find out the three plot tests of country, keyword, and institution.

4. Design/Methodology/Approach

The data was gathered from the webserver of Clarivate analytics, popularly known as Web of Science (WoS), a platform used by a wide range of scientific studies in bibliometrics. Wang et al.(Wang, Pan, Ke, Wang, & Wei, 2014) and Gorraiz and Schloegl (Gorraiz & Schloegl, 2008). The most comprehensive scientific research website (the ISI System of Science Website) mostly covered all the necessary data regarding the bibliometric analysis. We mainly used the Web of Science core collection database to cover more than 20,300 journals, books, and conferences with over 71 million records (AlRyalat, Malkawi & Monami, 2019).

A search has been carried out in the WoS database to get the overall results of the bibliometric publications. The query of searching is TS= "Information Retrieval"dated 28.05.2021. For the study, we have refined the search.DOCUMENT TYPE (Article OR Editorial Material OR Proceeding Paper OR Review OR Book Review OR Book Chapter OR Letter OR Review OR Early Access OR Meeting Abstract OR Correction) AND [excluding] DOCUMENT TYPES: (Data Paper OR Retracted Publication). Timespan: 2006-2020. Indexes: SCI-EXPANDED, SSCI, A&HCI. REFINED BY WEB OF SCIENCE. A total of 8206 documents were retrieved, 6915 of Articles, Proceeding Paper 853, Review 258, Editorial Material 99, Article; Early Access 50, Article; Book Chapter 27, Article; Data Paper 2 and Review and Book Chapter 2.

With a bibliometric analytical framework, the study explores the Web of Science core collection database covering published articles from2006 to 2020. The study employed the Biblioshiny app specially designed for the Bibliometrix R package, Histcite, and a robust Java-based application ScientoPY, to convert, process, visualize and evaluate the selected dataset. Moreover, we utilized another java based apps VOSViewer, to provide some more complete visualization.

5. Result and Discussion

Based on the results of 8206 documents published by 1344 sources (journals, books, etc.), written by 17060 authors,708 single-authored documents, and 16352 multi-authored documents affiliated with 4693 institutions, and 106 countries. These documents received 210377 total citations. Presented an overview of the research in the IR field with the information related to the publication and citations trend by year, type of documents, language-wise distribution, most productive countries, organizations, and authors, preferred

types of sources of researchers, citations, and use of influential research; top-ranked papers, most frequently used author keywords; co-occurrence network in IR research, Trend Topics and Topic Dendrogram, of each word in IR literature, Collaboration Network (Author, Institutions and Country) andThree plot tests of Country, Keyword and Institutionbased on collected data.

5.1 Year-Wise Research Growth Trend

Figure 1 shows the year-wise frequency of publications and citations published from 2006 to 2020. The trend indicates that publication and citation have not gradually increased. The total number of publications is consistently published in IR research, but the number of sources was decreased. The trend shows that 2006-2020 were average publications in that period and produced 2010 marvelous as in that year's highest number of publications. In the case citation, we observed that in 2008 highest number of sources was made, and after that, citations were gradually decreased.

5.2 Documents Types of Publications

Figure 2 shows the type of documents published under the IRresearch area. It can be noted that out of 8206 research output, a total of 6915 of the publication published in the form of the Article, Article; Proceeding Paper 853, Review 258, Editorial Material 99, Article; Early Access 50, Article; Book Chapter 27,Article; Data Paper 2, Review and Book Chapter 2. It is observed that researchers prefer journals to publish and communicate their research out in the form of articles.

5.3 Distribution of Publication by Language

Figure 3 shows the Language-wise distribution of publications on IR research. It is found that the maximum of the research publications is written in the English language (7930) followed by Spanish (115), Portuguese (72), Chinese (35), French (22), German (13), Czech (7). The remaining publications are published in other languages such as Dutch, Japanese, Afrikaans, Croatian, Lithuanian, Polish, and Russian.

5.4 Influential Countries Associated with IR.

The top ten (10) highly publishing countries on IR literature were as presented in Table 1. The result shows that the USA is at the top of the list with 1772 publications, 42216 citations, 23.82 citation impact, and it received the highest 3828 total link strength.China is on 2nd rank with 1281 publications, 13528 citations, 10.56 citation impact, and 2225 total link strength. After this, England occupied the third position, with 619 publications, 10236 citations, 16.58 citation impact, and it received 1704 total link strength. Taiwan and Australiaare

at the bottom of the list, with 297 and 322 publications, respectively. It shows that the country Canada has received the highest citation impact (24.27).

5.5 Relevant Affiliation

The top ten (10) organizations were producing research publications on IR literature given in Table 2. It shows that the Chinese Academy of Sciences (China) is on the top of the list with 110 publications, 1752 citations, 15.94 citation impact, and total link strength of 39. University of Granada (Spain) on 2nd rank with 76 publications, 1839 citations, 24.04 citation impact, and total link strength 22.The University of Sheffield(England) on 3rd position with 70 publications, 1423 citations, 20.33 citation impact, and total link strength 93.RMIT University(Australia)and the University of Maryland(USA)are at the bottom of the list, with 47 and 48 publications. Observed that the University of Amsterdam (USA) had received the highest citation impact (46.39).

5.6 Influential Authors

Table 3 highlights the top ten (10) most prolific authors on IR presented with total publications, total citations, Citation Impact, H Index, G Index, and Publication year start. The results show that majority of the authors starting their publication in the year 2006. The list of most prolific authors shows that Zhang Y. is the most productive author with 37 publications, 269 citations, 7.27 citation impact, 9 H Index, 15 G Index. The author Zhang J.listed 2nd rank with 31 publications, 784 citations, 25.29 citation impact, 8 H Index, 28 G Index followed by Li H. with 30publications, 873 citations, 29.10 citation impact, and 16 H, 29 G Index. It was noticed that Li H. has the highest total citations, Citation Impact, H Index, and G Index than the total listed authors.Fernandez-Luna J.M. is on the bottom of the list with 20publications, 72 citations, 3.60 citation impact, 5 H Index, 7 G Index.

5.7 Journal Wise Contribution to Publications

The journal's impact regarding the number of publications, citations, citation impacts, H Index, G Index, and starting publication year is highlighted in Table 4. It shows the journal " Information Processing & Management" is a highly influential journal producing a maximum of 367 publications, 6322 citations, 17.23 citation impacts,37 H Index, and 60 G Index. The "Journal of the American Society for Information Science and Technology" is on 2nd rank with 200 publications, 4697 citations, 23.49 citation impacts, 36 H Index, and 58 G Index; followed by "Expert Systems with Applications" with 175 publications, 4064 citations, 23.22 citation impacts, 30 H Index and 56 G Index. The "Online Information Review " is at the bottom of the list and has produced 102 publications, 1121 citations, 10.99 citation impacts, 16 H Index, and 29 G Index.

5.8 Highly-Cited Literature Analysis of IR.

The bibliographic information of the top ten (10) most cited articles is indicated in Table 5. The paper entitled "Image retrieval: Ideas, influences, and trends of the new age" by Datta R., published in 2008 in “ACM Computing Surveys," is on the top of the list with 1564 citations and 111.71 total citations per year. The article entitled “Hierarchical Dirichlet Processes" by The YW published in 2006 in "Journal of the American Statistical Association" is on 2nd rank with 1299 citations and 81.89 total citations per year. The article entitled "Content-based multimedia information retrieval: State of the art and challenges" by Lew M.S. in 2006 in "ACM Transactions on Multimedia Computing” is on 3rd rank with 787 citations and 49.19 total citations per year. It is noted that “Black holes as mirrors: quantum information in random subsystems.” is the bottom of this list, written by Hayden P.published in 2007 in “Journal of High Energy Physics” with 457 citations and 30.47 total citations per year.

5.9 Co-Occurrence Network of Author Keywords

Frequently used authors' keywords in IR research are highlighted in Figure 4. The keywords analysis has been performed in VOSViewer software. The minimum number of 5 keywords occurrence is selected, and hence only 813 keywords meet the threshold out of a total of 16114 keywords. The distance and size of the bubble indicate the number of keyword occurrences and associational links. 'Information Retrieval' is the most frequently used and representative author keyword that appears 2286 times and has 4445 total link strength, followed by 'Natural Language Processing' that appear 194 times and 440 total link strength, followed by 'Text Mining' that appears 165 times and 415 total link strength, followed by 'Machine Learning' that appear 163 times and 395 total link strength.

5.10 Term Analysis of all Keywords

Frequently used all keywords in IR research are highlighted in Figure 5. The minimum number of 5 keywords occurrence is selected, and hence only 1442 keywords meet the threshold out of 20371 keywords. The term 'Information Retrieval' is the most frequently and representative keyword as it appears 2286 times and 923 total link strength, followed by ‘Information-Retrieval’ that appear 727 times and 417 total link strength; ‘Model’ that appear 350 times and 345 total link strength, ‘Web’ that appear 333 times and 302 total link strength.

5.11 Trend Topics in IR.

Trend topic of author keywords in IR research is highlighted in Figure 6. The most frequent keywords in the last 15 years to observe the latest trends in Information Retrieval research. Topic trends are also part of this research, shows an overview of the development of the topic from time to time with the

division per year. It is known what topics have been used for a long time and what topics have been used recently. The emergence of topics adjusted to the frequency of the number of words appearing in research on Information Retrieval. The figure shows that 'Information Retrieval' 896 frequency is the most trending word in the year 2013, 'Retrieval' 121 frequency (2013), 'Information' 105 frequency (2015), 'Evaluation 85 frequency (2012), 'Machine Learning' 76 frequency (2014) are the top five keywords that are repeated most frequently in Information Retrieval literature from 2006 to 2020. 'Private Information Retrieval,' 'Semantics,' 'Deep Learning,' 'Capacity' was the most trending keyword in 2019. 'Privacy' is the only trending keyword in the year 2020.

5.12 Clustering and Scaling of Key Concepts of IR.

For providing a structured representation of how the patterns of the themes are spreading within the dimension of integrated pest management by the multiscaling algorithm of Biblioshiny, we initiated the analysis by exploratory cluster analysis of the themes and sequences popularly known; as "'Topic Dendrogram."'The topic dendrogram tree diagram showing the most widely used topics and their relation to other topics and classification of these topics depicted in different colors and the relationship between the keywords generated by hierarchical clustering. Figure 7 is showing a Topic Dendrogram of the top 50 author keywords of IR literature. The result shows that there are two major topic clusters. Cluster 1 consists of five (5) keywords 'Experimentation', 'Performance', 'Theory', 'Design' and 'Algorithms'. Cluster 2 consists of eight (8) sub-clusters and one single keyword, 'Future Extraction,' where each sub-cluster consists of sub-clusters. Sub-cluster 2.1 consists of certain topics on 'Retrieval' and 'Information.' Sub-cluster 2.2 consists of certain topics on 'Search' and 'Recommender System.' Sub-cluster 2.3 consists of certain topics on 'Information Retrieval', 'Digital Libraries', 'Web', 'Search Engines', 'Internet' and 'Text'. Sub-cluster 2.4 consists of certain topics on 'Evaluation', 'Ranking', 'Query', 'Clustering', 'Learning', 'Query-Expansion', 'Classification' and 'Search Engine'. Sub-cluster 2.5 consists of certain topics on Relevance 'Feedback' and 'Personalization.' Sub-cluster 2.6 consists of certain topics on 'Music', 'Music Information Retrieval', 'Similarity', 'Semantic Similarity', 'Geographic Information Retrieval', 'Ontologies', 'Image Retrieval', 'Privacy', 'Private Information Retrieval'. Sub-cluster 2.7 consists of certain topics on'Information Extraction', 'Semantics', 'Analysis', 'Databases', 'Learning to Rank', 'Data' and 'Text Mining'. Sub-cluster 2.8 consists of certain topics on 'Natural Language Processing,' 'Ontology,' 'Semantic Web,' 'Machine Learning,' 'Deep Learning,' 'Question-Answering,' 'Data Mining,' and 'Knowledge Management.

5.13 Author Collaboration Network of IR Research

The Collaboration Network or Collaboration between authors on IR research is presented in Figure 8. In this figure total of 50, author names are displayed,

some have a connection, and some are not. The authors' relationship is shown by clusters of color equations and lines between one term and another. The size of each square also indicates the number of papers published in this area. The figure shows the collaboration between the thirty-two (32) clusters of authors, but there are the two (2) most significant clusters in this study. The first cluster shows the collaboration between Wang J, Li Q, Lin HF, Lin Y, and Xu B. The second cluster shows Liu TY, Xu J, Wang I, and Li H. The authors who are not related and indexed in the data above show no collaboration between the author and other authors in making papers about the area of IR literature.

5.14 Institution Collaboration Network on IR Research

The Collaboration Network or Collaboration between Institutions on IR research is presented in Figure 9. In this figure total of 50 Institutions' names are displayed, some have a connection, and some are not. The figure shows the collaboration between the seventeen (17) clusters of institutions, but there are the two (2) most significant clusters in this study. The first cluster shows the collaboration between Wuhan univ, univ Massachusetts, hong kong polytech univ, Dalian univ technol, Nanjing Univ, Univ of Wisconsin, Nanyang technol univ, Chinese univ hong kong and city univ hong kong. The second cluster shows the collaboration between univ Chinese acad sci, Purdue univ, Beihang univ, Zhejiang univ, Fudan Univ, Tsinghua univ, univ elect sci and technol china, Rutgers state univ, natl univ Singapore, univ Illinois, and Carnegie Mellon univ.

5.15 Country Collaboration Network on IR Research

The Collaboration Network or Collaboration between Countries on IR research is presented in Figure 10. In this figure, we observed that 50 countries' names are displayed, and all countries connect. It is noted that the collaboration between the three (3) clusters of countries, but there is one (1) largest cluster in this study. The cluster shows the collaboration between Slovenia, Germany, Brazil, United Kingdom, Poland, Spain, Austria, Sweden, Belgium, Netherlands, Italy, Greece, Switzerland, Portugal, South Africa, Finland, Norway, Mexico, Argentina, Colombia, Chile, Romania, Denmark, and Cuba.

5.16 Three Plot Tests of Country, Keyword, and Institution

The three-factor diagram that has been generated of the top 10 keywords, countries, and institutions on literature IRresearch is highlighted in Figure 11. The size of the block shows the associational relationship with each factor. Figure 12 portrayed a more precise overview ofhow institutions from the various world had possessed scientific collaboration within specific keywords.The first plot represented the country, the second plot denoted the associated keywords, and the third plothighlighted the institutions working on

those keywords. The curve line stated the connections and the sizeof the plots comprised by the density of appearances.

6. Conclusions

This study's prime goal is to use the bibliometric approach for mapping the field of IR to examine and define the crucial literary works on this subject based on previous, ongoing, and future trends within this innovative field of research. This study employs bibliometric methodologies to analyze IR documents critically and provides a comprehensive summary of IR theories that develop overtimes. The analysis's foundation mainly comprised a visual overview of growth patterns, core research institutions, key researchers, key themes, and focus points in the IR study. Moreover, the evidence about the use of metadata diagrams and critical review based on the research network framework in such literature was found most profound, quantifying the exceptional novelty of the current study. The main research content and core themes of IR were further furnished with a coordinated strategic diagram that reflected the paper's core innovation.

The current research review used the bibliometrics method and visualization technology to analyze the literature on IR research indexed in the Web of Science during 2006-2020. Bibliometric analysis software packages, i.e., Biblioshiny, Histcite, and VOS-viewer, are used for data processing and extraction as bibliometric indicators. Most document types in this field are in the form of articles. There are 8206 documents published by 1344 sources (journals, books, etc.), written by 17060 authors,708 single-authored documents, and 16352 multi-authored documents affiliated with 4693 institutions and 106 countries. These documents received 210377 total citations. The USA is at the top of the list with 1772 publications, 42216 citations, 23.82 citation impact, and it received the highest 3828 total link strength. China is on 2nd rank with 1281 publications, 13528 citations, 10.56 citation impact, and 2225 total link strength. The Chinese Academy of Sciences (China) is on the top of the list with 110 publications, 1752 citations, 15.94 citation impact, and total link strength of 39. University of Granada (Spain) on 2nd rank with 76 publications, 1839 citations, 24.04 citation impact, and complete link strength 22. The author, namely Zhang Y., is the most productive author with 37 publications, 269 citations, 7.27 citation impact, 9 H Index, 15 G Index. Zhang J. listed 2nd rank with 31 publications, 784 citations, 25.29 citation impact, 8 H Index, 28 G Index. The journal " Information Processing & Management" is a highly influential journal producing a maximum of 367 publications, 6322 citations, 17.23 citation impact, 37 H Index, and 60 G Index. The "Journal of the American Society for Information Science and Technology" is on 2nd rank with 200 publications, 4697 citations, 23.49 citation impacts, 36 H Index, and 58 G Index. The article entitled "Image retrieval: Ideas, influences, and trends of the new age" by Datta R., published in 2008 in "ACM Computing Surveys," is on the top of the list with 1564 citations and 111.71 total citations per year. The article entitled "Hierarchical Dirichlet Processes" by The YW published in 2006 in "Journal of

the American Statistical Association" is on 2nd rank with 1299 citations and 81.89 total citations per year. 'Information Retrieval' is the most frequently and representative author keyword as it appears 2286 times and 4445 total link strength, followed by 'Natural Language Processing' that appear 194 times and 440 total link strength, followed by 'Text Mining' that appears 165 times and 415 total link strength. In the case of trends topics, we observed that 'Information Retrieval' appears highest numbers of time. 'Private Information Retrieval,' 'Semantics,' 'Deep Learning,' 'Capacity' was the most trending keyword in 2019. 'Privacy' is the only trending keyword in the year 2020. These trend topics indicate those IR researchers are now working on a web or digital environment. The collaboration between authors, institutions, and countries on Information Retrieval research results show some have a connection, and some do not. Several collaboration clusters show that many authors, institutions, and countries are collaborating in IR research.

It provides a comprehensive overview of publication trends on IR. It maps the key areas of this research field like the most productive countries, institutions, authors, journals, research categories, hotspots, and future research directions. It also enables the visualization of countries' collaborations in the IR field for partnership opportunities.The study represents the keywords picked by the authors within this research domain of IR, which will assess the future researcherto understand the key terms used by the IR researcher.These results are helpful for researchers to identify primary sources of publication for relevant information, thereby helping them improve their research direction and keep up with the research IR.

7. References

AlRyalat, S.A.S., Malkawi, L.W. & Momani, S.M. (2019). Comparing bibliometric analysis using PubMed, Scopus, and Web of Science databases. *Journal of Visualized Experiments*

Bar-Ilan, J. (2017). Bibliometrics of "information retrieval" - A tale of three databases. *CEUR Workshop Proceedings*, *1888*, 83–90.

Bar-Ilan, J., John, M., Koopman, R., Wang, S., Mayr, P., Scharnhorst, A., & Wolfram, D. (2016). Bibliometrics and information retrieval: Creating knowledge through research synergies. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–4. https://doi.org/10.1002/pra2.2016.14505301023

Ding, Y., Chowdhury, G. G., Foo, S., & Qian, W. (2000). Bibliometric Information Retrieval System (BIPS): A web search interface utilizing bibliometric research results. *Journal of the American Society for Information Science and Technology*, *51*(13), 1190–1204. https://doi.org/10.1002/1097-4571(2000)9999:9999<::aid-asi1031>3.0.co;2-b

Ding, Y., Foo, S., & Chowdhury, G. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *International Information and Library Review*, *30*(4), 367–376. https://doi.org/10.1080/10572317.1998.10762484

Glänzel, W. (2015). Bibliometrics-aided retrieval: where information retrieval meets scientometrics. *Scientometrics*, *102*(3), 2215–2222. https://doi.org/10.1007/s11192-014-1480-7

Gorraiz, J., & Schloegl, C. (2008). A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science. *Journal of Information Science*, *34*(5), 715–725. https://doi.org/10.1177/0165551507086991

Quoniam, L., Balme, F., Rostaing, H., Giraud, E., & Dou, J. M. (1998). Bibliometric law used for information retrieval. *Scientometrics*, *41*(1–2), 83–91. https://doi.org/10.1007/BF02457969

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

Roshdi, A., & Roohparvar, A. (2015). Review: Information Retrieval Techniques and Applications. *International Journal of Computer Networks and Communications Security*, *3*(9), 373–377. Retrieved from www.ijcncs.org

Safder, I., & Hassan, S. U. (2019). Bibliometric-enhanced information retrieval: a novel deep feature engineering approach for algorithm searching from full-text publications. *Scientometrics*, *119*(1), 257–277. https://doi.org/10.1007/s11192-019-03025-y

Saracevic, T. (1999). Information Science. *Journal of American Society for Information Science*.

Sheikhshoaei, F. (2020). Mapping Global Knowledge Domain, Research in Information Retrieval in Medical Sciences : A Scientometric and Evaluative Study. *Research Square*, 1–17. https://doi.org/https://doi.org/10.21203/rs.3.rs-184663/v1

Wang, B., Pan, S. Y., Ke, R. Y., Wang, K., & Wei, Y. M. (2014). An overview of climate change vulnerability: A bibliometric analysis based on Web of Science database. *Natural Hazards*, *74*(3), 1649–1666. https://doi.org/10.1007/s11069-014-1260-y

Ying Ding, G. G. . C. and S. F. (1993). Mapping the development in Information Retrieval specialty : A bibliometric analysis via journals. *Journal of Information Science*, *25*(1), 1–10. Retrieved from P143387632@ntu.edu.sg; Asggchowdhury@ntu.edu.sg; assfoo@ntu.edu.sg Abstract

Zhuhadar, L., Nasraoui, O., Wyatt, R., & Romero, E. (2009). *Evaluating Multi-Model (Metadata-Semantic) Information Retrieval System*.

Appendices

Tables

Table 1: Top Ten Influential Countries on IR

Country	Documents	Citations	Citation Impact	Total Link Strength
USA	1772	42216	23.82	3828
China	1281	13528	10.56	2225
England	619	10263	16.58	1704
Spain	533	7424	13.93	1030
Germany	389	6234	16.03	914
France	363	5454	15.02	837
Canada	362	8786	24.27	1047
India	329	1946	5.91	578
Australia	322	4934	15.32	805
Taiwan	297	3733	12.57	443

Table 2: Highly Productive Research Institutions

Organization	Country	TP	TC	CI	Total Link Strength
Chinese Academy of Sciences	China	110	1753	15.94	39
University of Granada	Spain	76	1839	24.20	22
The University of Sheffield	England	70	1423	20.33	93
Wuhan University	China	65	567	8.72	49
University of Illinois	USA	62	1556	25.10	25
Rutgers University	USA	54	1371	25.39	113
University of Amsterdam	Netherlands	51	2366	46.39	50
University of Wisconsin	USA	50	705	14.10	64
University of Maryland	USA	48	896	18.67	183
RMIT University	Australia	47	1358	28.89	58
TP- Total Publication, TC- Total Citation, CI- Citation Impact					

Table 3: Most Relevant Authors

Author	**TP.**	**TC**	**CI**	**H Index**	**G Index**	**PY Start**
Zhang Y.	37	269	7.27	9	15	2006
Zhang J.	31	784	25.29	8	28	2006
Li H.	30	873	29.10	16	29	2006
De Rijke M.	26	378	14.54	12	19	2006
Huang J.X.	22	139	6.32	8	10	2010
Li Q.	22	449	20.41	10	21	2006
Liu Y.	22	254	11.55	7	15	2006
Wang J.	22	528	24.00	9	22	2006
Kim J.	20	183	9.15	7	13	2007
Fernandez-Luna J.M.	20	72	3.60	5	7	2006
TP- Total Publication, TC- Total Citation, CI- Citation Impact, PY- Publication Year						

Table 4: Top Ten Highly Influential Research Journals

Source	**TP.**	**TC**	**CI**	**H Index**	**G Index**	**PY Start**
Information Processing & Management	367	6322	17.23	37	60	2006
Journal of the American Society for Information Science and Technology	200	4697	23.49	36	58	2006
Expert Systems with Applications	175	4064	23.22	30	56	2006
IEEE Access	145	438	3.02	9	15	2014
Multimedia Tools and Applications	137	873	6.37	15	23	2006
Information Retrieval	129	2426	18.81	27	44	2006
Journal of Documentation	127	1522	11.98	23	31	2006
Journal of the Association for Information Science and Technology	115	937	8.15	15	23	2014
Journal of Information Science	105	1100	10.48	14	26	2006
Online Information Review	102	1121	10.99	16	29	2006
TP- Total Publication, TC- Total Citation, CI- Citation Impact, PY- Publication Year						

Table 5: High-Cited Literature Analysis of IR.

Author	**Paper**	**Source Title**	**TC.**	**TC per Year**
Datta R.(2008)	Image retrieval: Ideas, influences, and trends of the new age.	ACM Computing Surveys	1564	111.71
Teh Y.W.(2006)	Hierarchical Dirichlet Processes.	Journal of the American Statistical Association	1299	81.19
Lew M.S.(2006)	Content-based multimedia information retrieval: State of the art and challenges.	ACM Transactions on Multimedia Computing	787	49.19
McGowan j.(2016)	PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement.	Journal of Clinical Epidemiology	607	101.17
Stamatatos E.(2009)	A survey of modern authorship attribution methods.	Journal of the American Society for Information Science and Technology	555	42.69
Salakhutdinov R.(2009)	Semantic hashing.	International Journal of Approximate Reasoning	554	42.62
Joachims T.(2009)	Cutting-plane training of structural SVMs.	Machine Learning	466	35.85
Zobel J.(2006)	Inverted files for text search engines.	ACM Computing Surveys	460	28.75
Zhou T.(2010)	Solving the apparent diversity-accuracy dilemma of recommender systems.	Proceedings of the National Academy of Sciences of the USA	458	38.17
Hayden P.(2007)	Black holes as mirrors: quantum information in random subsystems.	Journal of High Energy Physics	457	30.47
TC- Total Citation				

Figures

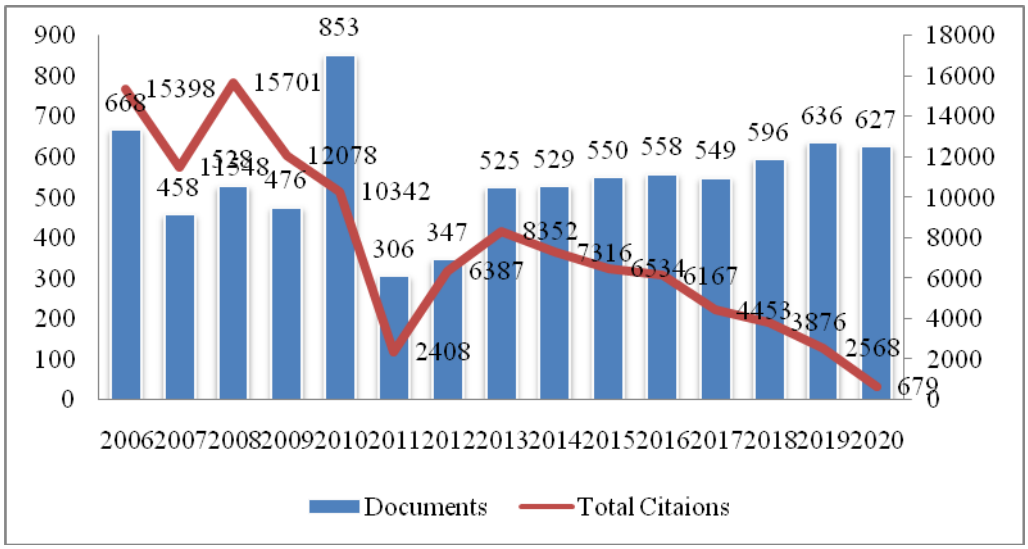

Figure 1: Publication and citations trend

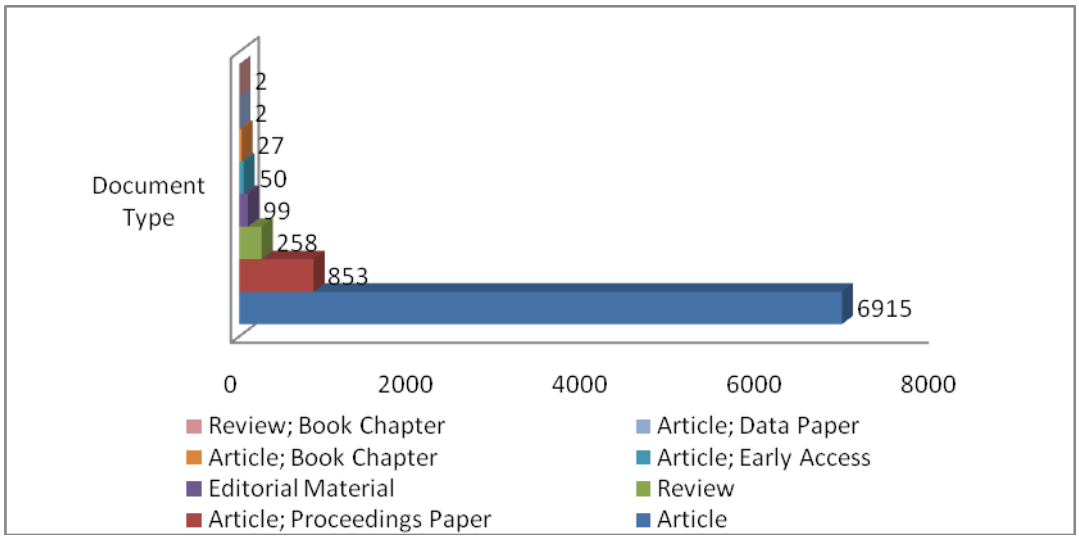

Figure 2: Type of documents

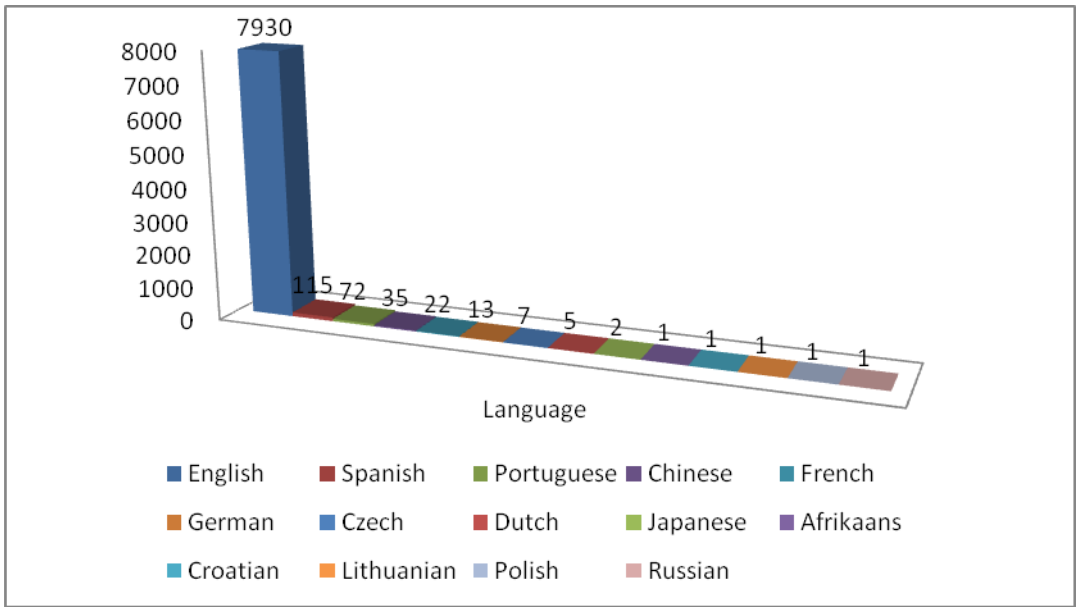

Figure 3: Language wise distribution

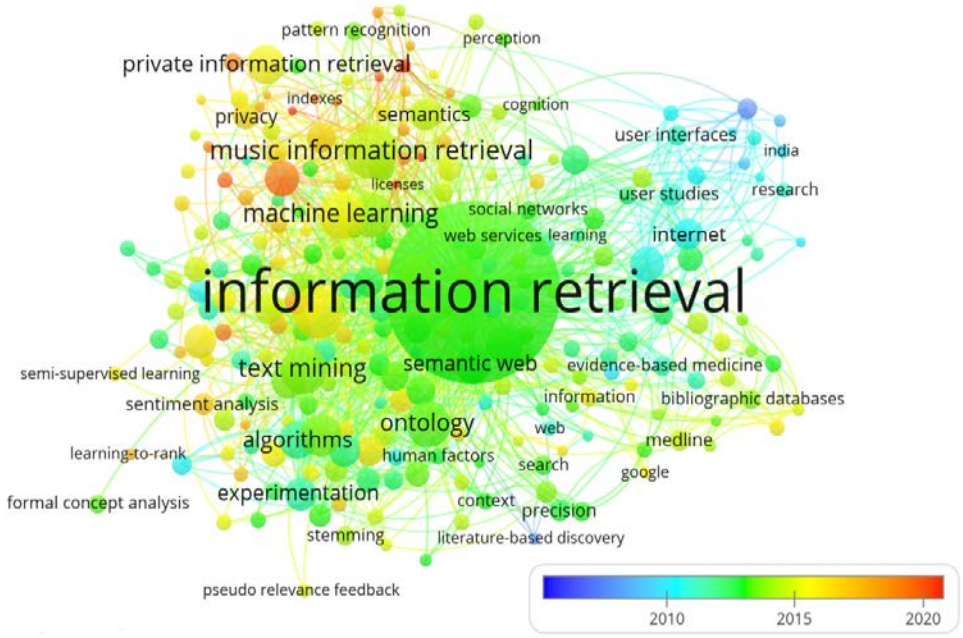

Figure 4: Co-Occurrence Network of Author Keywords

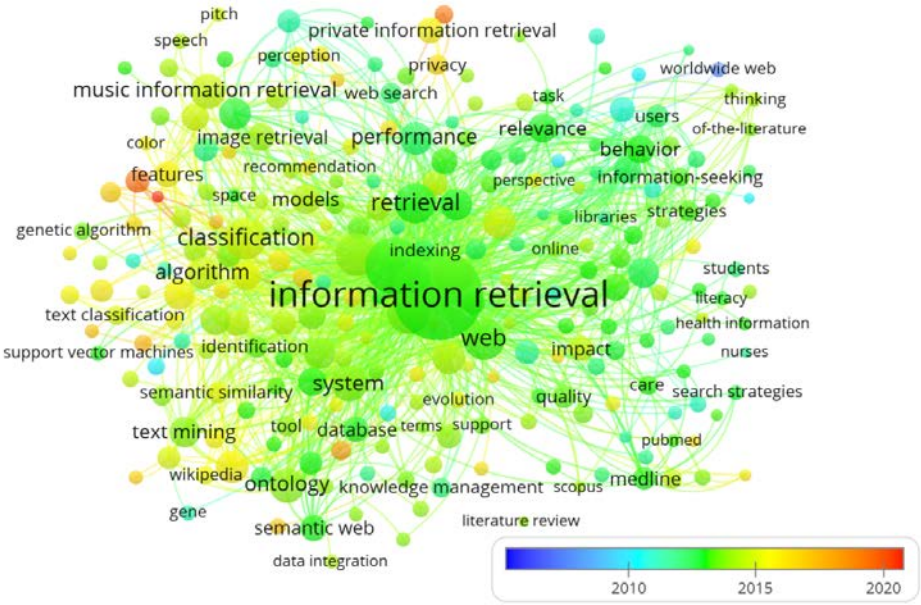

Figure 5: Term Analysis

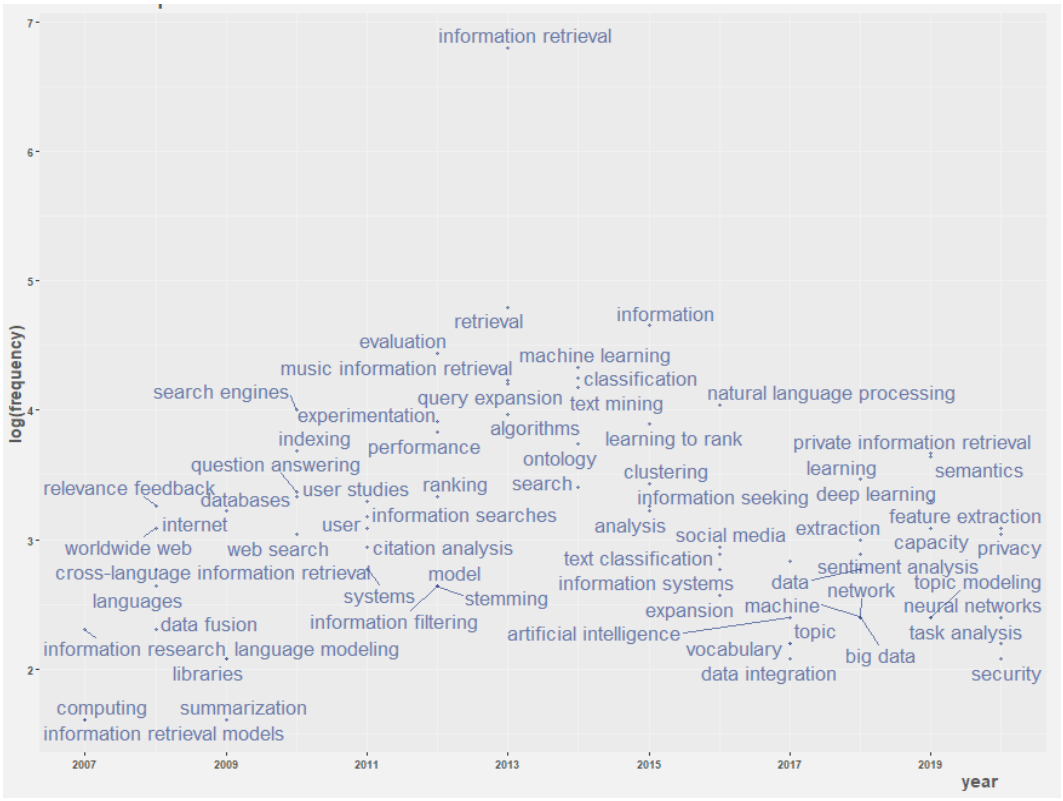

Figure 6: Trend topics on IR in 2006-2020

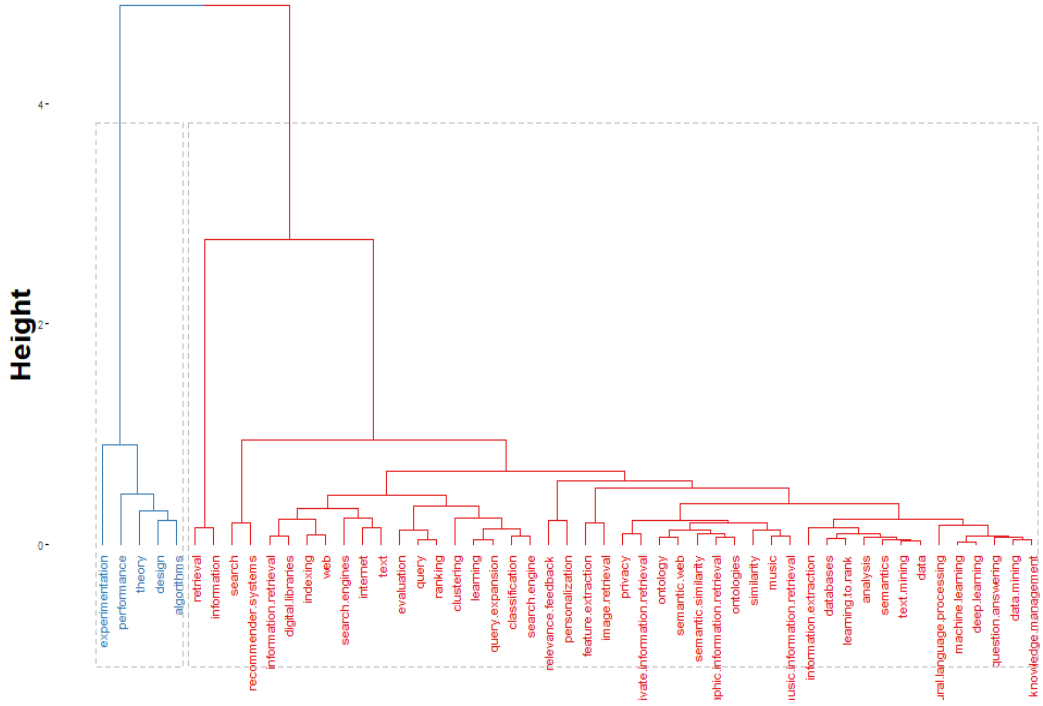

Figure 7: Clustering and Scaling of Key Concepts via Topic Dendrogram on IR

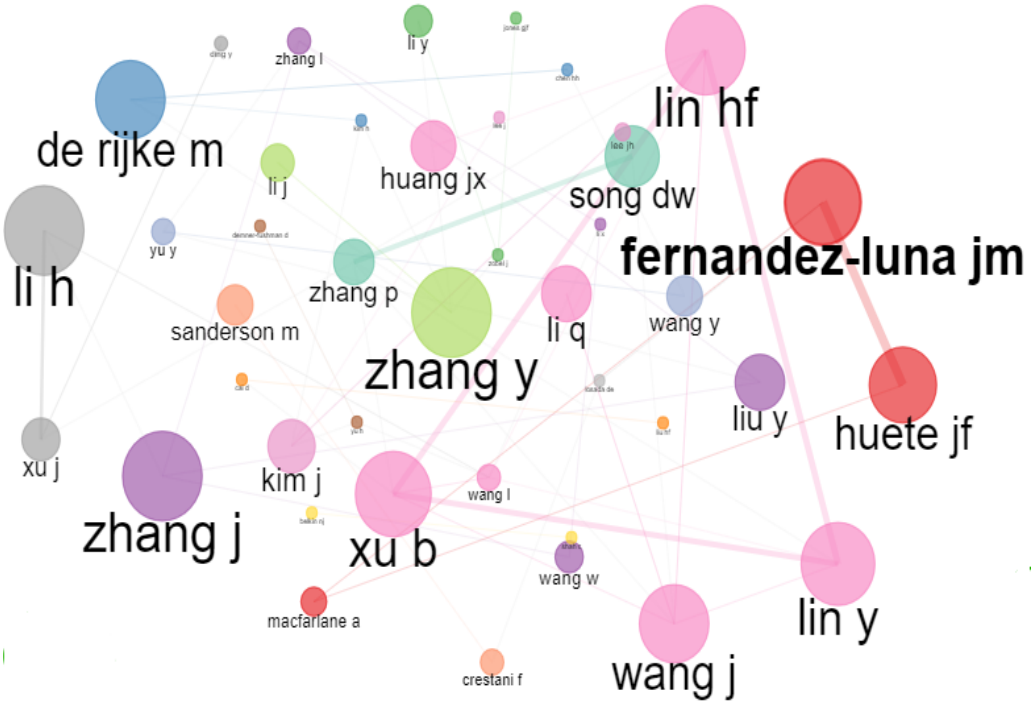

Figure 8: Authors Collaboration Network

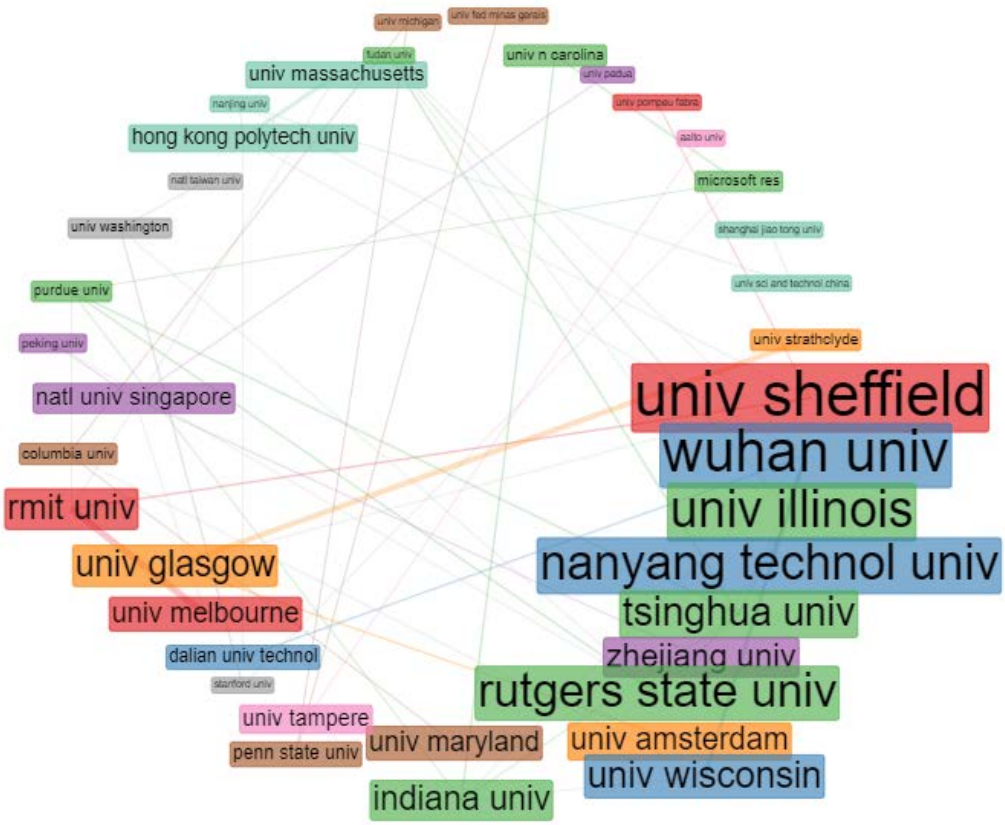

Figure 9: Institutions Collaboration Network

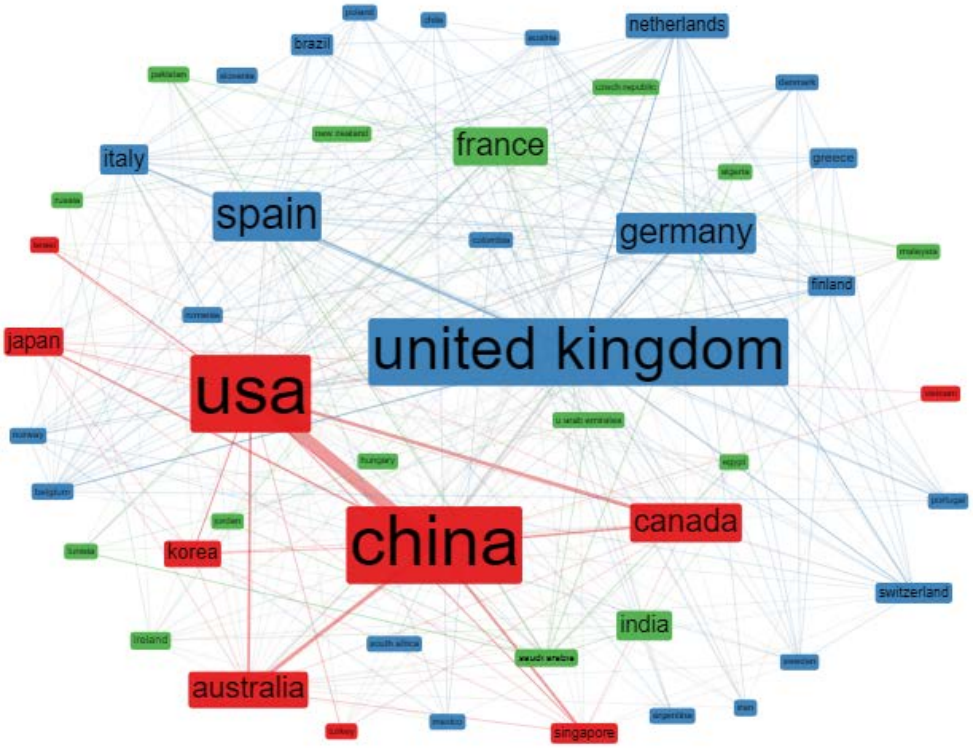

Figure 10: Country Collaboration Network

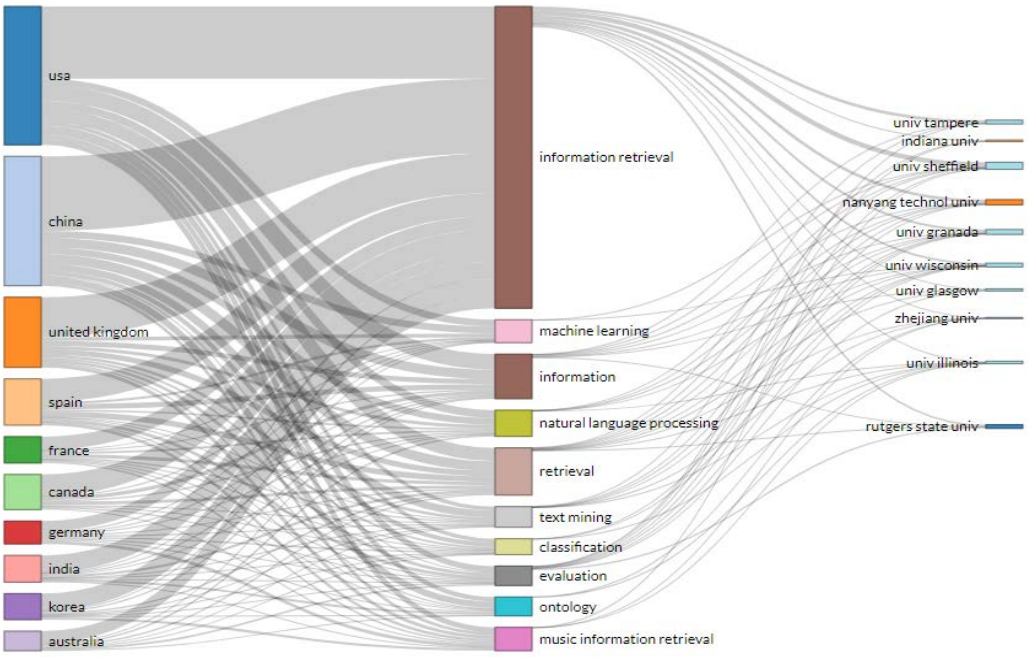

Figure 11: Three-Plot Analysis of Relationship among Country (left), the keyword (middle), and institution (right)