

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

2022

A Content Analysis Based Review on the Application of Document Similarity Measures on E-resource Retrieval in Agriculture and Allied fields

Nitumika Gogoi
nitumika.gogoi@aaau.ac.in

Uma Kanjilal
IGNOU, ukanjilal@ignou.ac.in

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Scholarly Communication Commons](#)

Gogoi, Nitumika and Kanjilal, Uma, "A Content Analysis Based Review on the Application of Document Similarity Measures on E-resource Retrieval in Agriculture and Allied fields" (2022). *Library Philosophy and Practice (e-journal)*. 7095.

<https://digitalcommons.unl.edu/libphilprac/7095>

A content analysis based review on the application of Document Similarity Measures on e-resource Retrieval in Agriculture and Allied fields

Mrs. Nitumika Gogoi
Research Scholar, IGNOU
Assistant Librarian, Assam Agricultural University, Jorhat
Email: nitumika.gogoi@aaau.ac.in (Corresponding author)

Prof. (Mrs) Uma Kanjilal, Pro Vice Chancellor
Deptt. of Library & Information Science
School of Social Science, IGNOU
Email: ukanjilal@ignou.ac.in

Abstract

From observations made by Robson (1993) who states that “the success of data collection should be assessed in relation to the specific research question”, it is felt necessary to study the aim as well as the fact that the most suitable unit of analysis will be sufficiently large to be considered as a whole but small enough to be referred as a relevant meaning unit during the analysis process (Graneheim & Lundman, 2004). Based on these criteria, the researcher tries to investigate that the study is carried out in the right direction which rest upon the credibility and essentiality of a content analysis of the literature accumulated for review.

Purpose: To measure the significance of literature reviews and the effectiveness of content analysis as analytic tool for the present research theme “Document Similarity Measures on e-resource Retrieval in agriculture and allied fields”. Content analysis is expected to enhance systematic literature reviews of various inductively derived sub-fields of Information Retrieval implemented for accessing information and data in Agriculture and allied fields.

Methodology: The present study is based on content analysis on reviews of 59 selected articles published between 1951 and 2020 that are retrieved after screening through the PRISMA software. Almost all related subtopics of Information Retrieval have been explored in the literature review compiled for the research topic. The analysis of the data is simultaneously supported with the data visualization by the application of the Tableau Desktop Public software.

Findings: Each document is assessed based on the contents of the documents collected for the study viz. year of publication, name of journals, number of authors, name of authors, research area, methodology, number of references cited in each paper, country to which author(s) belong. Collaborative research mostly conducted by authors with specialization in computer science and data science is seen. Documents compiled for the present study comprised mostly of comparative studies among the different Information Retrieval Models. Also it has been judged in several documents the effectiveness of Vector Space Model over the other IR models simultaneously verifying that content analysis is an imperative tool to

identify whether the literature review conducted so far is progressing in the appropriate direction.

Research Limitations: The major limitation is the limited number of documents retrieved and considered for the content analysis thereby customizing them for empirical analysis and assessing the significance of the literature reviews in order to identify the variations in the various research studies.

Conclusion: The analysis, software used and findings derived is aimed at facilitating the Information Retrieval researchers with the idea of conducting literature reviews through content analysis approach.

Keywords: Agriculture, Content analysis, Literature Review, Information Retrieval, Relevance effectiveness, Vector space model, PRISMA, Tableau software

Paper Type: Literature Review

Introduction

“You should write a review that leaves a clear impression of what is ‘well understood’ and what still remains a ‘mystery’ to be solved”

Diptak Bhattacharya, Researcher

Due to huge proliferation of studies worldwide, it has become difficult to keep up with the primary literature. Necessity arises for attempts to pull together and summarize on the specific topics of study. In any research document, a review of related literature is explicitly compiled into a chapter “Literature Review” as well as frequently included as a section in the introduction and the methodology chapters.

The subject “Information Retrieval” is basically categorized as applied research. But as studies on it progressed, many other studies are found to correlate with IR. These studies revolve around the fact called “retrieval effectiveness” which proves to be the basic scale of measurement of user satisfaction level and their information seeking pattern and need. As user study delineates the entire library functioning in an organization, so it has become very important for the information professionals to equip themselves with techniques of retrieving appropriate information demanded by the various categories of users. Today IR research is a conglomeration of various other research methods and the most common are the analytical, quantitative and experimental research methods. The output of IR research as has been suggested by Park (1993) results in relevance judgement which is treated as binary (relevant and non-relevant) or on a scale measuring degrees of relevance. Relevance judgements for evaluation of search results have been made either by search intermediaries or subject experts. Undoubtedly library professionals are actively playing the role of search intermediaries for the users.

Researchers have to strongly agree with the statement given by Seuring and Gold (2012) that “reviewing literature is an important supporting tool for other stages of the research process as well. Anchoring one’s first ideas in the existing body of literature will most probably be of great help for formulating and clarifying the research topic in the first place. Furthermore, unexpected insights from one’s own data analysis may induce the researcher to delve into literature at a later stage again in order to reconceptualise the findings.”

Review of related literature

Studies on literature reviews are considered as one of the demanding task in conducting a research. One of the universally acclaimed principle objective of reviewing literatures as stated by Hart (1998) is that literature reviews help in narrowing down the research topic as well as explain and justify research objectives, over all research design and methodology used. He states that novel research findings are discussed against the background of the existing body of literature, thereby confirming and rejecting, contrasting and complementing previous research outcomes. Literature reviews are also considered as scholarly contribution that map, consolidate and develop theory of a certain research area, thus facilitating subsequent research on the existing topic. Kitchenham (2004) defined systematic literature review as a means of identifying, evaluating and interpreting all available research relevant to a particular research question or phenomenon of interest and cited a few important reasons for undertaking a systematic review such as to summarise the existing evidence concerning a treatment or technology, identify any gaps in current research in order to suggest areas for further investigation as well as to provide a framework/background in order to appropriately position new research activities.

Seuring and Gold (2012) who cited Mentzer and Kahn (1995) and Meredith (1993) with reference to literature review that it provides valid tools for synthesizing and refining scattered knowledge regarding all stages of research conducted in a certain field (Mentzar and Kahn, 1995, cited by Seuring and Gold). This has been termed as the “normal research cycle” by Meredith (cited by Seuring and Gold) who conceived theory building as an ongoing iterative process running through the stages of description, explanation and testing of descriptive models and frameworks transformed into explanatory models which are then empirically tested that eventually develop as established theories. Seuring and Gold also stressed upon the fact that replication of the research and traceability of the arguments and conclusions call for more transparent and systematic procedures for conducting literature reviews and therefore content analysis offers methodological framework for conducting rigorous, systematic and reproducible literature reviews.

Berelson (1952) who has been cited in many research papers on content analysis defined the term as “a research technique for the objective, systematic and quantitative description of the manifest content of communication”. Berelson justifies “the process of analysis as liable and learnable method that precludes the personal authority of the researcher”. He stressed mainly on the concepts, techniques and objectives to be considered for content analysis of the subject matter. Contradicting Berelson’s definition on the fact that it failed to capture the qualitative and latent perspective of the analysis led to further investigation. With much discussion on the definition, Downe-Wambolt (1992) defined content analysis to be “more than a counting process, as the goal is to link the results to their context or to the environment in which they were produced”, it is rather “a research method that provides a systematic and objective means to make valid inferences from verbal, visual, or written data in order to describe and quantify specific phenomena”. Objectifying content analysis as applicable both in a quantitative and a qualitative approach without specifying the depth of analysis, Krippendorff (2004) redefined content analysis as “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use”. Further studies on the applicability of content analysis, introduced new thoughts like that of Yildirim and Simsek (2008) who understood that data can be defined as well as can be revealed of their meaning through content analysis because the basic process is to bring the similar data within the framework of certain concepts and themes together and organize them in a form that can be understood by readers. This was later understood clearly from Khakpour (2012) who stated that when the results from different articles or studies are compared, it would enable us to obtain a comprehensive understanding about the studies in a specific area or a specific concept. Later in the studies of Saban (2009) and Camnalbur, Bayraktar & Amuce (2013), it has been evaluated that content analysis is a widely used qualitative research method which includes screening printed or visual materials systematically and analyzing them based on identified categories thematically.

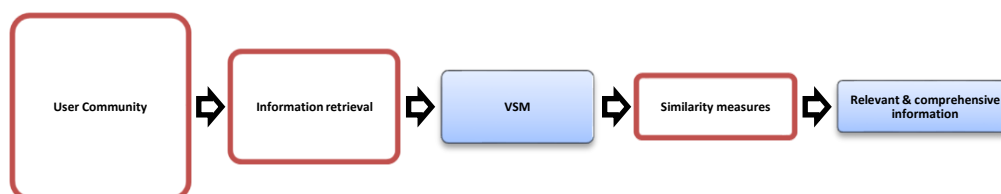
The development of Information Retrieval (IR) system witnesses the progressive journey from manual library operations of acquiring, indexing, disseminating and searching information to integrated automated library and information services. Therefore the conventional metaphor for information retrieval (a user submits queries to a static collection of information and retrieves a static collection of documents) is inadequate, and should be supplemented with an information retrieval technique able to take into account the dynamic changing nature of information stored in digital libraries. Other important technical challenges which will affect the traditional information retrieval process arise from the need for distributed storage, distributed retrieval (Croft, 1995), multimedia retrieval, richer interaction between users and their information sources and, iterative query refinement

(Ramana, 1995). Many studies so far and still undergoing, have tried to map IR in different context although the readings lead to similar understanding. It is only the implementation of IR that varies from content to content. Research studies, reports, theories and many different literatures on IR are created and disseminated on various online platforms. But need for specificity arises when it comes to studies on different aspects of IR. Therefore from the above studies it is realised that quality assessment on IR can be obtained through an appropriate content analysis. It is understood from the opinions of Elo & Kyngäs (2008), Harwood & Garry (2003) that content analysis focuses on a simplified presentation of the contents and results supplemented by figures and visualization and that the research phenomena earn a more conceptualized approach through the analysis process. However, at the same time, they equally contradict that conception may have different objectives since content analysis generally does not include a technique to connect concepts. They finally concluded that “the main consideration is to ensure that the structure of results is equivalent and answers the aim and research questions.”

Aims and objective

The aim of this paper is to highlight the significance of literature reviews and the effectiveness of content analysis as analytic tool for the present research theme “Document Similarity Measures on e-resource Retrieval in agriculture and allied fields”. Content analysis is expected to enhance systematic literature reviews of various inductively derived sub-fields of Information Retrieval implemented for accessing information and data in Agriculture and allied fields.

Specifically this paper focuses on reviewing IR based Vector Space Model and its application in retrieval of documents specific to agriculture and allied fields. It is objectified to provide a succinct overview of the whole research study undertaken with the title “*Application of Document Similarity Measures on e-resource retrieval in Agriculture and allied fields*”. The title itself can be partitioned into three different topics to have made in-depth studies: ‘*document similarity measures*’, ‘*e-resource retrieval*’ and ‘*agriculture and allied fields*’. The broad subject matter being Information Retrieval, it is knitted with concepts of information seeking, document relevance, ranking, similarity measures, precision and recall. It is imperative to advocate upon the three highlighted key aspects outlined in red in the following figure.



Scope

The present review is expected to

- Broaden the ideas of the application of Information Retrieval models with regards to information generation and indicates areas subjective to further research either in terms of user studies or manual query classification for librarians and information professionals as well as can be implied as a supported document to researchers of data science. Also discussions on major debates in IR, methodological and theoretical limitations can also be tracked from the literatures cited in this review.
- Support studies on user behaviour and attitude towards the search strategies adopted for retrieving relevant information
- Serve as a basic guide for studying the information seeking behaviour and challenges of the different categories of users of agricultural information and its allied fields.

Limitation

One of the primary limitations is that not all papers or book chapters from different online platform can be aggregated due to the vastness of literatures flooded over the web.

Research question

To draw a conclusion that this content analysis based review article is conducive to be referred to all kinds of Information Retrieval studies conducted in the field of library and information science.

Methodology and data collection

The present paper is based on the review of literatures collected on Information Retrieval and its components under the following queries from researcher's end:

1. Agriculture and IR
2. Importance of User groups and studies
3. Query collection and classification
4. Vector Space Model (VSM) and the similarity measures
5. Relevance : precision and recall

Criteria

- The first criterion for inclusion of papers was that at least two reviews of the respective sub-fields have been published in relevant peer-reviewed journals in the ten years.

- Secondly criteria for content analysis of the literature reviews on the topic presented are:
 1. Year of publication
 2. Name of journals
 3. Number of authors
 4. Name of authors
 5. Research area
 6. Methodology
 7. Number of references cited in each paper
 8. Country to which author(s) belong

Step-wise description

1. A systematic literature search is conducted using the following research terms: Information Retrieval System (IRS), Vector Space Model (VSM), relevance, term frequency (tf), inverse document frequency (idf), precision and recall, similarity measures.
2. Scientific journal papers from subject areas “mathematics” and “computer/data science”, social science journal papers from subject areas “agriculture” and “library and information science” as well as book chapters from these disciplines are retrieved from various online platforms specifically ResearchGate (social networking site), Scopus (Elsevier’s abstract and citation database), Citesheer (digital library for scientific and academic papers) and various open access journal sites.
3. Conducting a primary search, a total of 110 papers are identified and based on the backward reference search, backward author search, forward reference search and forward author search suggested by Levy and Ellis (2006) an additional search was conducted for relevant papers published in the past and future by the authors of the papers retrieved during the primary search.
4. Kitchenham and Charters (2007) identified individual studies contributing to a systematic review as primary studies while systematic review as a form a secondary study. Applying the PRISMA 2020 statement and the guidelines of Kitchenham and Charters, screening and quality assessment of the retrieved papers through primary search is conducted. Based on the review of the titles and abstract, 38 irrelevant papers as realised by the researcher are eliminated. A final quality assessment is again taken up which led to further elimination of 13 papers.
5. 59 papers were obtained after the screening and quality assessment was conducted. These papers are taken into account to determine scope of future research and derive the solution to the research question of the original study.

6. The number of papers obtained after the above filtration, were assessed and included in the content analysis is depicted by the PRISMA flow diagram generated from https://estech.shinyapps.io/prisma_flowdiagram/

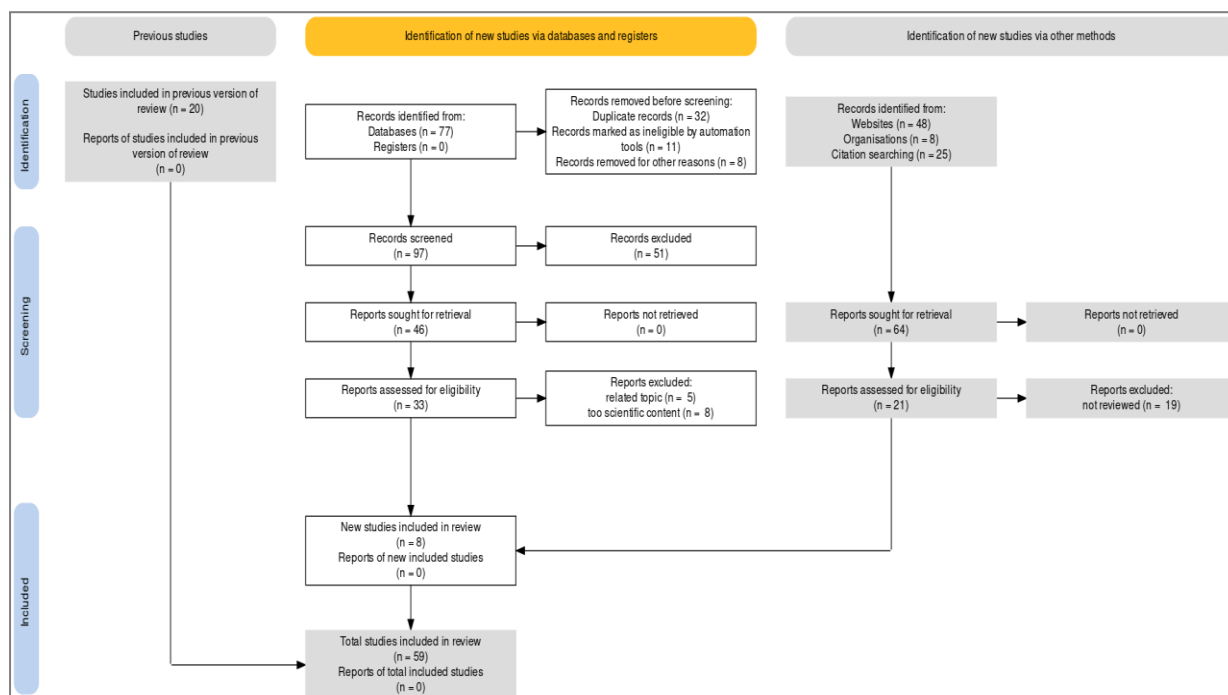


Figure 2. Selection of papers in the PRISMA flow diagram generated from https://estech.shinyapps.io/prisma_flowdiagram/

Analysis and Interpretation

Automated assistance is sought to carry out the analysis and correlations to facilitate the identification of data trends and relationships. Therefore along with statistical data representation of the Microsoft excel database the Tableau Desktop Public software is used to visualize the data. This helped to improve the recognition of the mutual patterns that exist in the literature by analyzing the number of papers per year, the journals in which they were published, number of citations in each articles that concentrates the depth of the study, number of publications per journal per year distribution of articles by approach, research types, research area and methodology used.

Methodology	No. of articles	%
Analytical	12	20.33
comparative	6	10.16
correlation	7	11.86
descriptive	7	11.86
experimental	7	11.86
evaluative	4	6.77
quantitative	11	18.64
schematic	5	8.47

Table 1: Article distribution based on methodolog

Table 1 illustrates that articles/books including almost all types of methodologies are included in the “Literature Review” on the research topic and it is analysed that maximum articles are analytical (20.33%) and as Information Retrieval incorporates the information needs of the users for retrieval effectiveness, literatures with quantitative methods occupies the second highest position (18.64%) in the review study.

In any kind of user specific study, the statistical method lets users describe their need and demands. Although experimental, descriptive, correlation and comparative research are different types of quantitative research methods, these are mentioned distinctly and only the statistical implementation in the various articles are represented as quantitative research method.

Publication year(class interval)	No.of articles(frequency)	%
1951-1960	4	6.77
1961-1970	0	0
1971-1980	2	3.38
1981-1990	4	6.77
1991-2000	10	16.94
2001-2010	13	22.03
2011-2020	26	44.06

Table 2: Article distribution based on publication years

Table 2 shows the trend of articles published from 1951 to 2020. It is found that the publication of articles gradually increases with the advancement of time. During 1951-1960, Information Retrieval being in its primitive stage, attracted good number of researchers to investigate on the topic. However during 1961-1970, not much work is noticed to have been done. Gradually from 1971, studies on Information Retrieval (IR) took pace and was conducted extensively and from the data gathered it can be concluded that proliferations in IR research drastically increased and different areas of IR viz. vector space model, Boolean model, probabilistic model, query classification, similarity measures/coefficient precision and recall with respect to document collection and user needs, are being continuously explored through different perspectives of application of IR. During the last decade, increase in IR research is visibly high (44.06%) and is actively studied as a specialization of the data science and computer applications with further research in correlation with digital library that has been interpreted from the content of the articles included in the study.

Research area	No. of articles	%
Information retrieval	18	30.5
Information seeking	2	3.38
Information science	5	8.47
Vector space model	10	16.94
Probabilistic model	1	1.69
Boolean model	1	1.69
Query classification	4	6.77
Similarity measure	10	16.94
relevance	6	10.16
User study	1	1.69
Digital library	1	1.69

Table 3: Article distribution by research area

In table 3, the literature reviewed includes articles mostly on IR (30.5%) as the topic of research itself is a component of IR. Moreover, the table shows that almost all aspects of IR have been discussed in the literature review to get a good understanding of the concept and then specifically focuses on the main areas upon which the research is based i.e. vector space model (VSM), query classification, similarity measures, and relevance. Among them, article content with VSM and similarity measures based article are consulted at length (16.94 % each), followed by relevance in the third position (10.16 %).

Ref. cited (class interval)	No. of articles (frequency)	%
0-29	42	71.18
30-59	11	18.64
60-89	3	5.08
90-119	1	1.69
120-149	1	1.69
150 and above	1	1.69

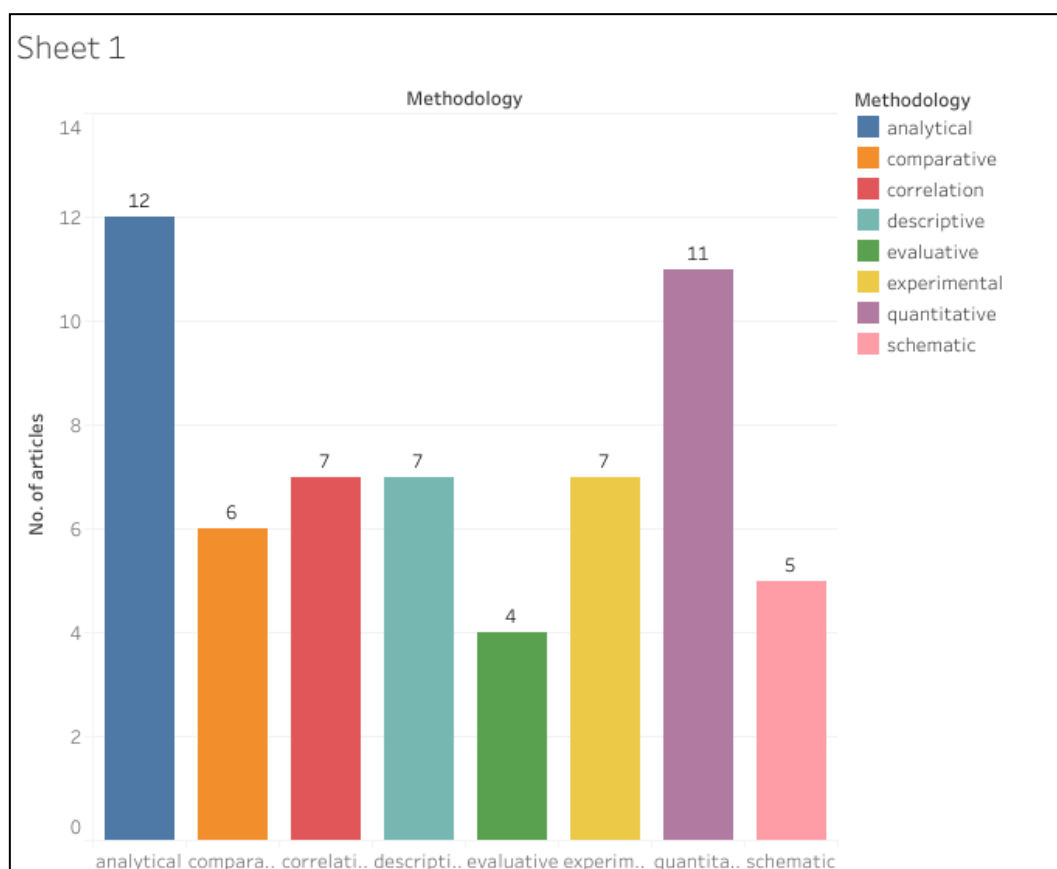
Table 4: Article distribution by cited references

A total of 6 documents are indicated as **NA** (not available) under the reference cited column as no data related to references could be extracted. However while analysing number of article with cited references the count is included under the range (0-29) in table 4. The highest number of references found in the articles is below 29 references (71.18%)

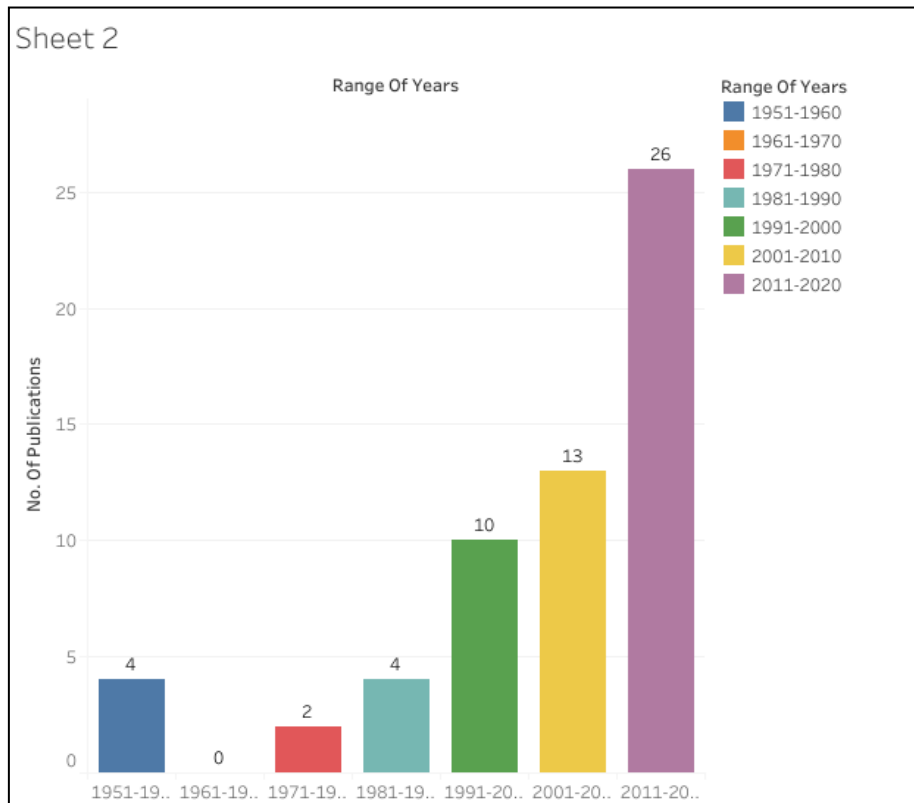
Only three articles with exhaustive cited references of 104, 134 and 231 conducted for research areas of information seeking, relevance and information science respectively are reviewed in this literature review.

Data visualization through Tableau Public Software version 2021.4

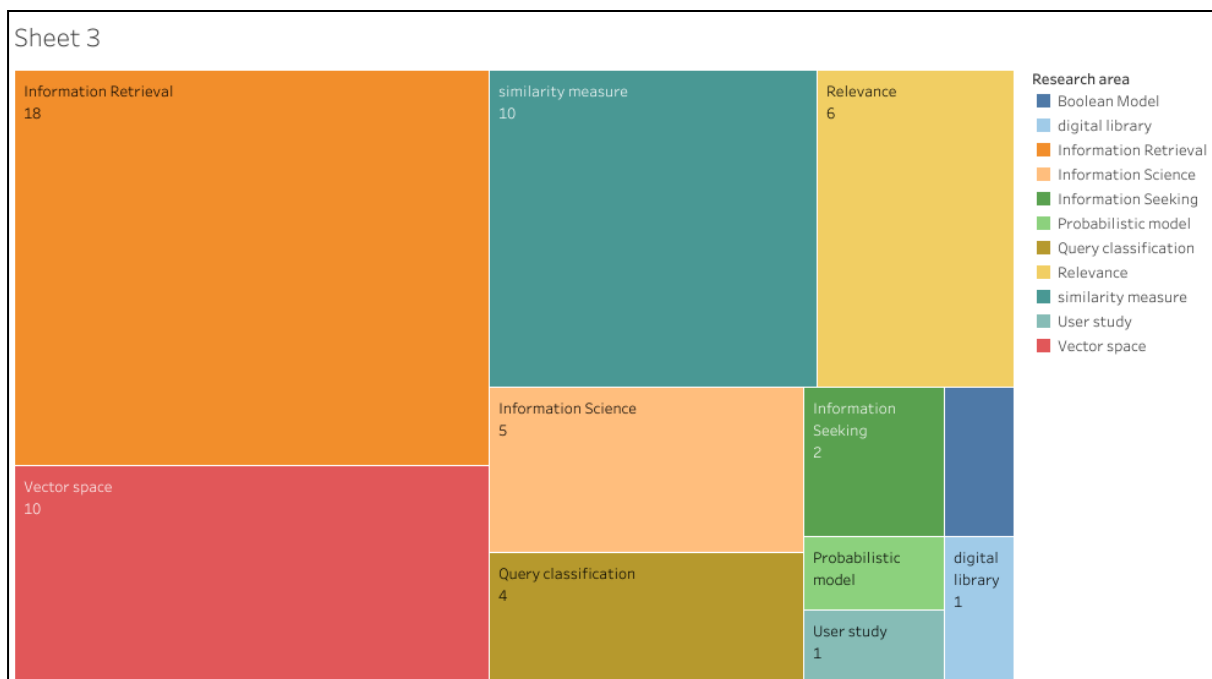
Sheet 1 is about the distribution of the number of articles based on methodology and is a visualization of the data in table 1.



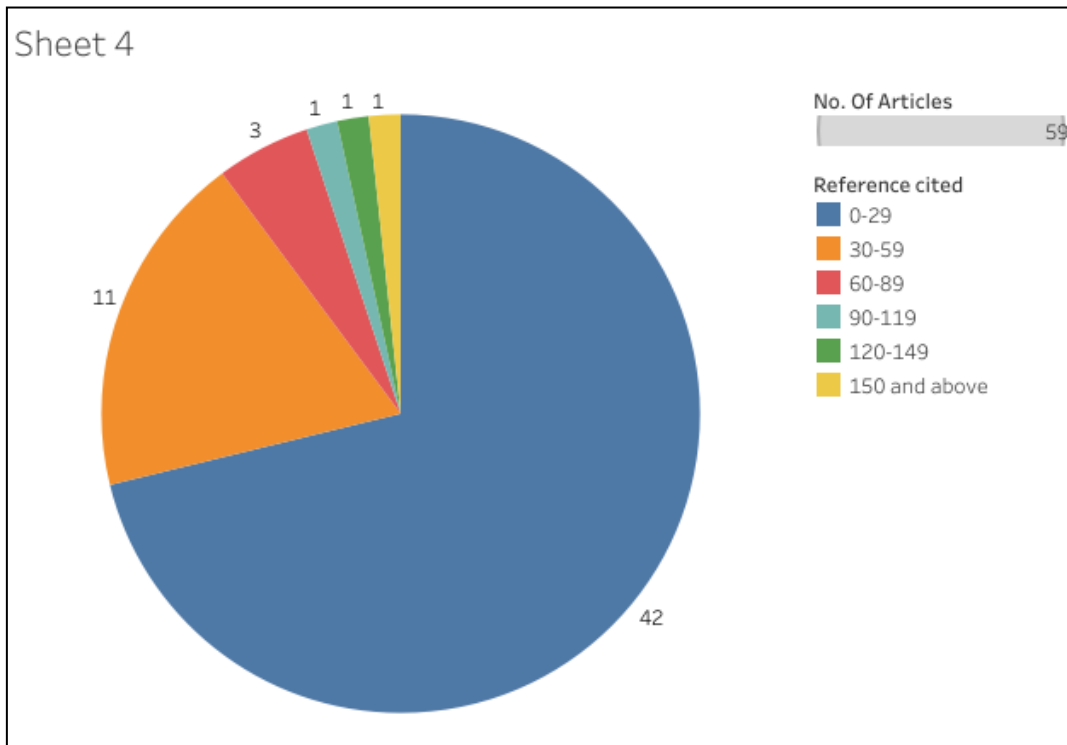
Sheet 2 represents the distribution of number of articles considered for the study published during 1951 to 2020 and it presents the pictorial representation of the publication trend of 59 articles in increasing order with the highest number of publications found during 2011 to 2020. This has been statistically depicted in table 2.



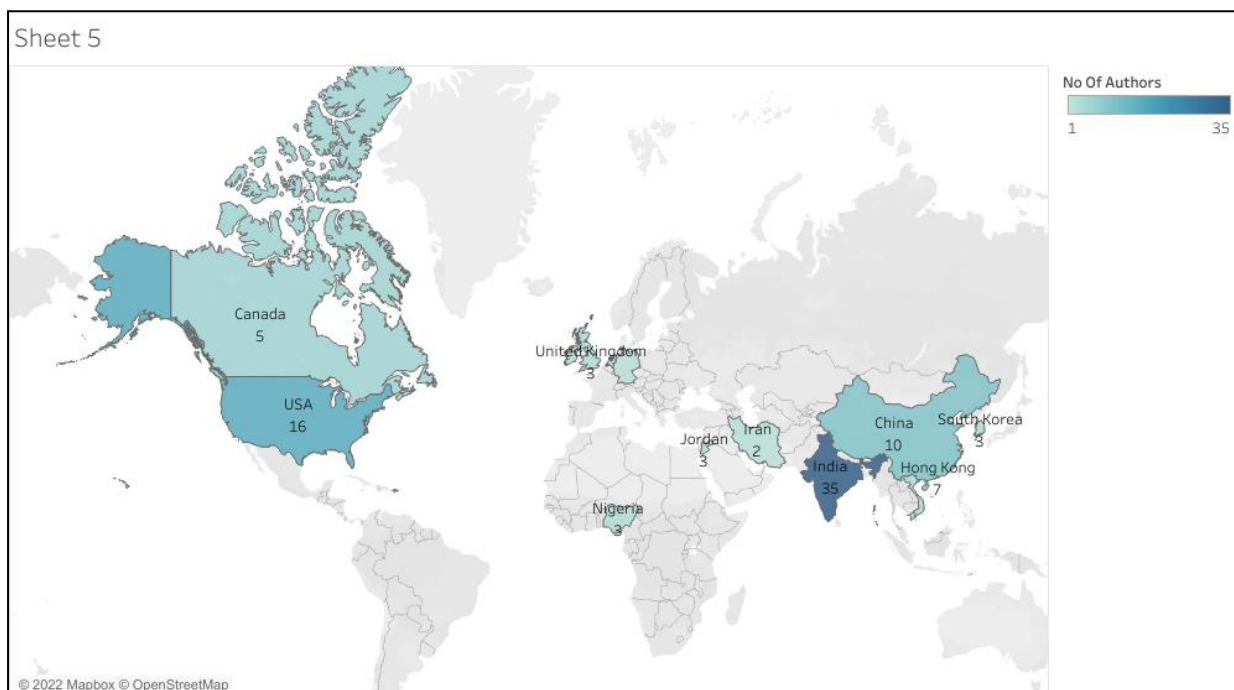
Sheet 3 is the visualization of the research areas focused for which the literatures are collected and it mainly highlights the areas of IR, VSM and Similarity measures.



Sheet 4 represents the pie diagram that takes into account the references cited in the articles. The highest number of references is found to be below 29 against 42 articles. The interpretation for table 4 is enhanced by the pictorial representation of sheet 4.



Sheet 5 represents the distribution of authors based on the geographic location and gives a view of the global scenario. The studies on IR and its component are widely spread out not only among the leading nations but nations like Nigeria, Iran and Jordan as well. The researcher however prefers to find out the degree of research studies conducted at the national level and therefore maximum of the article are concentrated at the Indian region (35)



Observation and result

It is noticed from the present study that maximum of the articles were studies done by authors belonging to computer science and data science subject background. Only a few were stated

to be information/library professionals. Therefore although IR takes a major role in the digital library system, the subject matter “digital library” is gradually becoming the cup of tea for the data scientists. In this regard effective studies need to be conducted by the library science fraternity as it is they who deem it the responsibility of information generation, storage, processing and dissemination.

Secondly the studies included in this review article indicate collaborative research in the field of IR as a number of studies with two/three authors seem predominant. It is therefore suggested that future studies should also be collaborative research centric for more comprehensive and objective results.

Thirdly as is already mentioned that most of the authors are from data science background, the articles were studies comprising the analysis of Cranfield datasets (Cleverdon, 1960) which has been the foundation of many IR studies conducted yet.

Assuming to arrive at the result to the research question framed and based on the above observations, the researcher strongly recommends that similar methods when adopted, the information professionals can attempt studies on Vector space model to evaluate the relevance factor between queries submitted by the users and the document corpus in the databases. The only necessary criterion is an exhaustive literature review followed by developing expertise in empirical, methodological and cognitive skills.

Conclusion

Although the content analysis tool is used to analyse the effectiveness of the literature collected for the study concerned, it cannot be judged to provide a specific outcome about their impact. It entirely relies upon the researcher how well he/she supports the inclusion of the literatures with meaningful statements w.r.t. the study as well as move from description and patterns to interpretation, determine the underlying meanings of concepts and relationships identified, and observe the gaps in the process, something that software is incapable of (Vidmar, et al, 2021). Fidel (1993) has technically explained in the words of Ellis (1992) that IR research can be defined and described in two paradigms viz physical and cognitive. Physical paradigm is the traditional experiment well demonstrated by Cranfield studies which still is referred and applied for various types of IR research. Although technically a successful study but it increases discontent among the researchers as they fail to explore the cognitive paradigm by ignoring the researcher-respondent relationship which is necessary to gain the degree of retrieval effectiveness. While physical paradigm relies on artefacts, cognitive paradigm demands for user study whereas IR research cannot be conducted in the absence of any of the two paradigms. As such these studies require both methodological and observation skills and the integrity to correlate them. Patton (1990) as

cited by Fidel therefore recommends that “generating useful and credible qualitative findings through observation, interviewing and content analysis requires knowledge, discipline, training, practice, creativity and handwork.”

Reference

1. Bengtsson, M. (2016). “How to plan and perform a qualitative study using content analysis.” *NursingPlusOpen* 2, 8–14 Elsevier
2. Berelson, B.L. (1952). *Content analysis in communications research*. New York:Free press.
3. Camnalbur, M., Bayraktar, M. D., & Amuce, E. N. (2013). The effect of web-based instruction on achievement: A meta-analysis study. *Cypriot Journal of Educational Sciences*, 8, (3), 292-301.
4. Cleverdon, C.W. (1960). The Aslib Cranfield Research Project on the Comprehensive Research Project on the Comparative Efficiency of Indexing Systems. *ASLIB Proceedings*. Emerald. 12(12):421-431. doi: 10.1108/eb049778
5. Conclin J. (1987) Hypertext: An Introduction and Survey. *IEEE Computer*, 20(9), 17-41.
6. Croft W. B. (1995).What Do People Want from Information Retrieval. *D-lib Magazine*
7. Debons, A., Horne, E., & Croenweth, S. (1988). *Information Science, an integrated view*. Boston: G.K.Hall.
8. Demiroka, M.S., Baglamaa,B., & Besgula, M. (2015). A Content Analysis of the studies in Special Education Area. *Procedia - Social and Behavioral Sciences*, 197, 2459 – 2467 Elsevier
9. Downe-Wamboldt, B. (1992). Content analysis: Method, applications and issues. *Health Care for Women International*, 13, 313-321.
10. Elo, S., & Kyngäs, H. (2007). The qualitative content analysis process. *J. Adv. Nurs*, 62 (1), 107–115, doi: 10.1111/j.1365-2648.2007.04569.x .
11. Elo, S., Kääriäinen, M., Outi Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open*, 1-10. doi: 10.1177/2158244014522633
12. Fidel, R. (1993). Qualitative methods in Information Retrieval Research. *Library and Information Science Research*, 15, 219-247.

13. Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, 24, 105-112.
14. Haddaway, N.R., Pritchard, C.C., & McGuinness, L.A. (2021). *PRISMA2020: R package and ShinyApp for producing PRISMA 2020 compliant flow diagrams*. Zenodo. doi :10.5281/zenodo.4287834
15. Hart, C. (1998). *Doing a Literature Review – Releasing the Social Science Research Imagination*, Sage Publications, London.
16. Harwood, T., & Garry, T. (2003). An overview of content analysis. *The Marketing Review*, 3(4), 479-498. doi: 10.1362/146934703771910080
17. Kang, J., & Gao, J. (2012). Application of Ontology Technology in Agricultural Information Retrieval. *Proceedings of the 2012 2nd International Conference on Computer and Information Application (ICCIA 2012)*, 1183-1186.
18. Khakpour, A. (2012). Methodology of comparative studies in education. *Contemporary Educational Researches Journal*, 1, 20-26.
19. Khirfan, L., Peck, M., & Mohtat, N. (2020). Systematic content analysis: A combined method to analyze the literature on the daylighting (de-culverting) of urban streams. *MethodsX*, 7, Elsevier. doi: 10.1016/j.scs.2020.102225
20. Kitchenham, B. (2004). *Procedures for Performing Systematic Reviews*, Keele University, Keele, UK.
21. Kitchenham, B., & Charters, S. (2007). Guidelines for Performing Systematic Literature Reviews in Software Engineering. *Software Engineering Group, School of Computer Science and Mathematics*, Keele University: Staffs, UK; *Department of Computer Science*, University of Durham: Durham, UK
22. Krippendorff, K. (2004). *Content Analysis: An Introduction to its Methods*, 2nd ed., Beverly Hills, USA SAGE
23. Levy, Y., & Ellis, T. (2006). A Systems Approach to conduct an Effective Literature Review in support of Information Systems Research. *Information Science*, 9, 181-212.
24. Luhn, H.P. (1958) The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2, 159-165. doi:10.1147/rd.22.0159
25. Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
26. Park, T.K. (1993). The nature of relevance in information retrieval: an empirical study. *Library Quarterly*, 63, 318-351.

27. Qi-ming, WU. (2010). Agricultural Information Retrieval System based on Comprehensive Information Theory. *Agricultural Science and Technology*, 11(2), 143-145.
28. Ramana R., Pedersen, J.O., Hearst, M.A. et al.(1995). Rich Interaction in the Digital Library. *Communications of the ACM*, 38(4), 29-39.
29. Robson, C. (1993). *Real world research: A resource for social scientists and practitioner-researchers*. Oxford, UK: Blackwell.
30. Salamphasis, M, Tait, J, and Bloor, C. (1996). *Cooperative Information Retrieval in Digital Libraries*. Citeseer.
31. Seuring, S., and Gold, S. (2012). Conducting content-analysis based literature reviews in supply chain management, *Supply Chain Management: An International Journal*, 17 (5), 544 – 555. doi: 10.1108/13598541211258609
32. Tableau Desktop, www.tableau.com
33. Vidmar, D., Marolt, M., & Pucihar, A. (2021). Information Technology for Business Sustainability: A Literature Review with Automated Content Analysis. *Sustainability*, 13, 1192, 1-24. doi: 10.3390/su13031192