IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Systematic Literature Review Of Tools And Techniques Used For The Construction Of Thesaurus

Nagare Nilesh V. (M.Lib.I.Sc.)
Research Scholar
Maharaja Sayajirao Gaikwad Arts, Science & Commerce College,
Malegaon Camp Dist. Nashik (Maharashtra)

Dr. Kale Vilas A. (Ph.D.)
Research Guide
Swatantrya Sainik Surybhanji Pawar College
Purna JN, Dist. Parbhani (Maharashtra)

Abstract:

A comprehensive look at the tools and methods used to create a thesaurus is provided in this paper. Its goal is to locate published works on the construction of thesauruses. This study employs a method known as a systematic review. Since 1960, a variety of information sources have been used to search the literature. 24 of the 152 research papers, conference papers, and book chapters were chosen based on specific inclusion and exclusion criteria. The study led to the discovery of previously unpublished research on the thesaurus-building tools and methods. Thesaurus structure and relationships, the role or significance of classification schemes in the construction of thesaurus, and other similar topics reviewed here are those used by various thesaurus developers in various fields. In the context of the creation of a thesaurus, the methodology and guidelines utilized in its construction are also examined in this study.

Keywords: thesaurus, thesaurus construction, classification, thesaurus relationships, descriptors, natural language, subject headings, storage and retrieval, Systematic literature review.

1. Introduction

Worldwide, knowledge is expanding on a daily basis. In this context, libraries and information centers play a crucial role by providing precise, comprehensive, and speedy information. The library and information industries have also seen a number of shifts as a result of the explosion of knowledge. The evolution of storage and retrieval systems is one of the most significant changes to libraries. Therefore, proper organization of information sources is essential for effective and simple information retrieval. The documents are arranged by the subject or subject content of the documents, in addition to the author and title. However, in the case of variety, the richness of natural language causes a number of issues with information storage when representing the subject matter of information sources. Through synonyms, homonyms, spelling variations, word forms, compound phrases, and other means, it causes issues. The precise retrieval of information from systems is influenced by this richness, or variety, of natural language.

As a result, it is necessary to control the terms that are used to represent the document's subject matter. In information storage and retrieval, various tools like a list of subject headings, classification schemes, and thesaurus are used to control such vocabularies. The literature on the subject is expanding at a very rapid rate in the context of the rapidly shifting scenario. It is becoming increasingly difficult to examine the accumulated evidence in this particular research area. As a result, research into vocabulary control tools like thesauruses is needed. According to Baumeister & Leary (1997)1, a literature review can assist in gathering and combining data from previous studies and determining the best course of action.

2. Methodology

A systematic review approach has been used in this study. A method for evaluating and interpreting previous research in a related field is known as a systematic literature review (SLR). In this study, the literature on various classification and thesaurus types, tools, and methods used in the past for the construction of thesaurus was searched from various databases and research journals in the field of library and information science. The literature was accessed online through research gate, N-LIST e-journals, e-books, and a variety of databases, as well as through data searches of printed journals and books.

3. Findings

Following paragraphs provide a review of selected literature dealing with

- the emergence of the concept of thesaurus
- the process of thesaurus construction,
- the role of classification and faceted classification in this context is also reviewed.

4. Review of literature

4.1 Thesaurus: Concept

In his 1960 article, Vickery, B. C.[2], emphasized the development of the thesaurus as a concept and as a tool for information retrieval. Using a variety of examples, the author demonstrated how a thesaurus can be a useful tool for information retrieval. He also talks about how HP Luhn made a dictionary of ideas by grouping words of notional families that are similar to and related to each other using the structure of Roget's Thesaurus. The author concluded by stressing the necessity of listing all keywords used in retrieval systems in a linear order.

The first widely accessible thesaurus for vocabulary control in an information retrieval system was published in 1960 by the Armed Services Technical Information Agency (U.S.)[3] under the name Thesaurus of ASTIA Descriptors 1960. In 1959, the Du Pont organization developed a predecessor of this kind for the same purpose.

Dr. Cyprian Kinner, who developed a unique classification system for botany, and George Dalgarno, an educator, who revised this classification system on the basis of Aristotelian categories, were the subjects of an investigation by Schulte-Albert, H. G. (1979)[4]. In contrast, Wilkins has developed a dictionary that functions similarly to a thesaurus and facilitates the display of a variety of relationships between individual terms.

A book on vocabulary control for information retrieval was written by Lancaster, F.W. (1986)[5]. The author of this book talked about aspects of vocabulary that are necessary for indexing and searching documents. Additionally mentioned are the factors that influence the performance of information retrieval systems, the organization of construction, and the display of vocabulary. From highly structured thesauri and classification schemes to searching with natural-language texts, a variety of vocabulary control tools are discussed. The author talked about the features and parts of an indexing language, the vocabulary that

the online retrieval system needs, and some of the benefits of vocabulary control tools that are cheap. The book is primarily intended for library and information science students.

4.2 Tools and techniques for the construction of thesaurus

Peterson (1990)[6] describes the procedures and policies used to create an indexing language that would represent art and architecture knowledge and serve as a substitute for images and objects. The "Atomized" presentation of AATs (Art & Architecture Thesaurus) and faceted language are also discussed by the authors. For free text searching, most bibliographic databases offer access to descriptors from a controlled vocabulary and text words. Fidel, R. (1992)[7] studied 47 professional searchers who worked in various types of libraries and specialized in a variety of subject areas to observe online search practices. The significance of controlled vocabulary is demonstrated by this study. Searchers consulted thesauri for 75% of the selected search terms, demonstrating the significance of thesauri and indexing. According to the study, the searcher who preferred text words did not use synonyms to make up for their lack of vocabulary control. Additionally, this study demonstrates that quality searching necessitates the use of both text words and descriptors.

Spiteri, L. F. (1998)[8] looked at a number of different national and international guidelines for making thesaurus. According to the study, these guidelines emphasize the various approaches to the construction of alphabetical thesauri but provide very little instruction on how to use facet analysis in the thesaurus. After the Information Retrieval Thesaurus of Education Terms was published in 1968, facet analysis was first used to build faceted thesauri. Due to the growing development of online information systems and the use of free text as a medium to retrieve the necessary information from them, Aitchison, J., Bawden, and A. Gilchrist (2000)[9] appropriately emphasized the necessity of a thesaurus. However, the precise information might not be retrieved from the free text that is accessible in the natural language of words in the title, abstract, and full text of papers. Therefore, vocabulary control tools like thesaurus, subject headings, and classification systems are necessary. One of the most useful resources for finding precise information is the thesaurus. They have provided in-depth explanations of the construction and use of the thesaurus.

Through a selected bibliography, Nielsen, M. L. (2004)[10] focused on the issues and issues in the construction of thesaurus and suggested some readings that might be useful for thesaurus design. The author has provided a list of articles that have discussed the thesaurus's concept, purpose, and evolution in response to new information retrieval technologies over time. The problems and progress in the collection and

organization of thesaurus concepts and terms, as well as the various methods used in their construction, are covered in detail in these articles.

The process of creating a thesaurus of library and information science terms using a speciator-based faceted depth classification schedule was described by R. Kumbhar (2005)[11]. Additionally, he discussed the advantages of using this strategy to construct various thesaural relationships in the LIS thesaurus. The characteristics, benefits, and drawbacks of thesauri and other forms of controlled vocabulary in electronic format that can be found on the internet were discussed by Laguens Garcia, J.L. (2006)[12] The author demonstrated the classifications and thesaurus available online with the help of examples.

An automated method for adding new terms to thesauri as entry vocabulary was described by Wang, Jun (2006)[13]. It can be added by looking at how words and phrases in titles and subject descriptors are linked to one another. The author employed methods for steps like

1. Obtaining words and phrases from the metadata's title field.
2. To select relevant and specific keywords from a thesaurus used to catalog the document using associated controlled vocabulary terms.
3. To include the chosen keywords as new terms in the thesaurus. As a result, it is simple to update the thesaurus with new terms used in literature.

Broughton Vanda (2010)[14] talked about how to use Bliss Bibliographic Classification to create a thesaurus for law from a faceted classification. The author demonstrated the classification's structure and methods for creating the sonic relationships. Broughton also provided examples to illustrate the design process for the alphabetical index and schedule display. A comprehensive guide for the creation and management of monolingual controlled vocabularies was provided by the National Information Standards Organization (2010)[15]. It focuses on controlled vocabularies that are used to represent content in thesauri and other knowledge organization systems.

4.3 Classification in construction of thesaurus

Using an analytical synthetic faceted method of classification developed by Dr. S.R. Ranganathan, Ghosh, S., and Panigrahi, P. (2015)[16] developed the ontology for the field of library and information science. With the assistance of the seventh edition of Dr. S.R. Ranganathan's Colon Classification Scheme, the hierarchical relationships between the terms collected for the construction of the thesaurus have been established. The

canon of characteristics and its associated canons have been utilized in the process of defining the hierarchy of class-subclass among terms. The significance of Dr. S.R. Ranganthan's philosophy has been demonstrated by this work.

The various patterns of classification systems and the introduction of depth classification patterns were described in Kaula (1985)[17]. The author also observes and discusses the limitations of faceted classification systems as well as their universal applicability. Kaula proposes the reconstruction or depth version of schedules and thesaurus as an alternative for the development of classification in the future for information retrieval.

The value of classification in general cognitive skills was established by Gopinath (1986)[18], who also published the instructions for creating depth versions of colon classification. From the kind of resources required to draft a depth classification schedule to an example of a depth version of colon classification for solar energy, Gopinath further described the entire process or step of designing a depth classification schedule. Facet analysis, first proposed in colon classification by Singh, Sukhdev (1999)[19], is now widely accepted. In addition, he refers to the analytic synthetic pattern of classification as a technique for subject analysis and its application into constituent aspects based on the fundamental categories and other subject features of the schemes.

According to Foskelt, D.J. (2001)[20] the detailed order of colon classification probably outperforms any other method. It achieves two major goals in library classification by providing an explicit facet formula for each class.

a. A helpful order so that the material is always arranged in accordance with the most important aspects of the subject in each class.

b. It is almost certain where a simple compound or complex subject is located.

In order to construct the thesaurus of library and information science, Kumbhar (2004)[21] created a depth classification schedule for the field. Dr. S.R. Ranganathan's speculative and pragmatic approaches were combined to create this schedule. The author followed Gopinath M. A.'s guidelines when creating the depth classification schedule. Broughton, V. (2006)[22] conducted a study to estimate the impact of faceted classification and the faceted analytical method on the creation of various information retrieval tools. According to the study, faceted systems have become more common in recent decades. The majority of library and information science subject indexing tools, including classifications, subject heading lists, and

thesauri, now include facet analysis features. The author provided an overview of the information retrieval methodology's faceted analytical approach as well as its various considerations and applications.

In developing the knowledge organization system, Asundi, A. Y. (2013)[23] has applied the principles of the helpful sequence that were provided by Dr. S.R. Ranganathan. He put a lot of emphasis on the connections that exist between helpful sequence principles and knowledge organization systems. The arrangement of subject isolates is heavily influenced by these principles. The knowledge organization system uses the "Principle of Later-in-Evolution" to sequence terms, and it can also be used to design thesaurus and taxonomies. A study by Spiteri, L. F. (2000)[24] looked at how many thesauri used the same facet analysis principles and which ones were used in the construction of thesauri. The measuring instrument for evaluating the structure of faceted thesauri was created by the author. The fourteen-faceted information retrieval thesauri were evaluated with it. It discovered that some thesauri arranged their indexing terms using both facet analysis and enumerative classification methods. While only half of the thesauri employ the synthesis principle, the number of facets used varies.

The development and nature of the Chinese Thesaurus (CT) were examined by Zhang, W. (2020)[25], and the author also evaluated the organizational structure of Chinese information indexing. The Chinese Thesaurus and the Chinese Classified Thesaurus (CCT), according to the study, have both contributed significantly to the standardization of Chinese language retrieval.

4. 5. Conclusions

Systematic reviews are useful for researchers looking for evidence for their study and are an important tool for summarizing evidence. The aforementioned study was looked at to find out what has been written about the tools and methods used to build a thesaurus. This systematic literature review revealed the following three main aspects: the development of the idea behind the thesaurus, the tools and methods used to build it, and the significance of faceted and depth classification in making connections between terms chosen for the thesaurus. We were able to comprehend the significant method that a number of authors and professionals working in libraries used to construct the thesaurus using the principles of helpful sequences by reviewing this literature. Dr. S.R. Ranganathan's analytico synthetic faceted method, which he used to create the schedules for the classification of the colon and the depth, is also useful for this purpose. The findings of this study will aid in future research on the topic.

5. References

1. Baumeister, R. F., & Leary, M. R. (1997). Writing Narrative Literature Reviews. Review of General Psychology, 1(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311
2. Vickery, B. C. (1960). Thesaurus - A New Word In Documentation. Journal of Documentation, 16(4), 181–189. https://doi.org/10.1108/eb026286
3. Armed Services Technical Information Agency (U.S.). (1960). Thesaurus of ASTIA descriptors. Washington, DC: United States Government Printing Office.
4. Schulte-Albert, H. G. (1979). Classificatory Thinking from Kinner to Wilkins: Classification and Thesaurus Construction, 1645-1668. The Library Quarterly, 49(1), 42–64. https://doi.org/10.1086/600900
5. Lancaster, F. W. (Frederick W. (1986). Vocabulary control for information retrieval. Information Resources Press.
6. Petersen, T. (1990). Developing a New Thesaurus for Art and Architecture. Library Trends, 38(4), 644–658.
7. FIDEL, R. (1992). Who need controlled vocabulary? In Special Libraries (Vol. 83, Issue 1, pp. 1–9).
8. Spiteri, L. F. (1998). The Use of Facet Analysis in Information Retrieval Thesauri: An Examination of Selected Guidelines for Thesaurus Construction. Cataloging & Classification Quarterly, 25(1), 21–37. https://doi.org/10.1300/J104v25n01_03
9. Aitchison, J., Bawden, D., & Gilchrist, A. (2000). Thesaurus construction and use : A practical manual. ProQuest Ebook Central http://ebookcentral.proquest.com
10. Nielsen, M. L. (2004). Thesaurus Construction: Key Issues and Selected Readings. Cataloging & Classification Quarterly, 37(3–4), 57–74. https://doi.org/10.1300/J104v37n03_05
11. Kumbhar R.M. (2005) Speciator based depth classifications application in thesaurus construction. Annals of library and information studies. 52(1) P.15-24
12. Laguens García, J. L. (2006). Thesauri and controlled vocabularies in Internet. Information Science Journal, 9(0 SE-Artículos), 105–121.
13. Wang, J. (2006). Automatic thesaurus development: Term extraction from title metadata. Journal of the American Society for Information Science and Technology, 57(7), 907–920. https://doi.org/10.1002/asi.20352

14. Broughton, V. (2010). The Use and Construction of Thesauri for Legal Documentation. Legal Information Management, 10(1), 35–42. https://doi.org/10.1017/s1472669610000289
15. National Information Standards Organization. (2010). Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. In NISO Press. National Information Standards Organization (NISO).
16. Ghosh, S., & Panigrahi, P. (2015). Use of Ranganathan's analytico-synthetic approach in developing a domain ontology in library and information science. In Annals of Library and Information Studies (Vol. 62).
17. . Kaula, P.N.(1985) A Treatise on Colon Classification New Delhi: Sterling Publishers Pvt. Ltd.. XV, 314p.
18. Gopinath, M.A. (1986) Construction of depth version of colon classification: a manual. New Delhi: Wiley Eastern Ltd.
19. Singh, S. (1999). Potentialities, limitations and the future of the Colon Classification. Herald of Library Science, 38(3/4), 186–193.
20. Foskett, D.J. (2001) Consecutive analysis of major classification scheme in practice some useful solution. IJILIS 14(1-2). P.32-38
21. Kumbhar,R.M. (2004) Thesaurus of library and information science. New Delhi : Ess Ess Publication P.172
22. Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. Aslib Proceedings, 58(1/2), 49–72. https://doi.org/10.1108/00012530610648671
23. Asundi, A. Y. (2013). Principles for Helpful Sequence and Deduction of Knowledge Organization Systems - An Exploratory Study. Journal of Information Science Theory and Practice, 1(2), 6–15. https://doi.org/10.1633/jistap.2013.1.2.1
24. Spiteri, L. F. (2000). The Essential Elements of Faceted Thesauri. Cataloging & Classification Quarterly, 28(4), 31–52. https://doi.org/10.1300/J104v28n04_05
25. Zhang, W. (2020). The development and structure of the Chinese Thesaurus for subject indexing. International Information & Library Review, 36, 47–54. https://doi.org/10.1080/10572317.2004.10762623
26. Aitchison J. Gilchrist,A and Bowden D.(2000) Thesaurus construction and use: manual 4th ed. London: ASLIB. P.99
27. Ranganathan, S. R. (1989) Prolegomena of library classification Vol.1 3rd Ed. Benglore: Sarada Ranganathan Endowment for Library Science P.124