CHAPTER – 2

THESAURUS CONSTRUCTION AND ITS ROLE IN INDEXING

CHAPTER - 2

THESAURUS CONSTRUCTIONS AND ITS ROLE IN INDEXING

2.1   Introduction

A thesaurus is a representation of keywords associated with a subject domain(s). It is also known as a tool for vocabulary control that guides indexers as well as end users to know the use of terms and helps in improving quality of search results. Generally, a thesaurus is designed for indexing and subsequent retrieval of documents in a specific subject domain. Examples of subject areas covered by thesauri are ERIC for education resources, macrothesaurus for economic resources, GESIS thesaurus for social science, legal thesaurus developed by R.L. Burton, etc,.

Thesaurus plays an important role in the organization of information and its subsequent retrieval. According to Aitchison et al. (2000), primary role of the thesaurus is information retrieval. Information retrieval is one of the most important aspects of librarianship as it deals with needs of end users of the library.

2.2   Composition of Thesaurus

A thesaurus provides different types of information for indexers as well as for end users. A thesaurus will comprise of terminologies which are used for describing terms and semantic relations with each other. These terms are preferred terms/descriptors, non-preferred terms/non-descriptors, related terms, narrow

terms(NT), broad terms(BT), USE and Used For(UF), etc,. For effective use of thesaurus while indexing documents and allocating appropriate keywords one has to have a clear understanding of relations in it. An attempt has been made to define all the relations used in a thesaurus.

2.2.1 Preferred Terms/Descriptor

It is obvious that a thesaurus needs to indicate what all terms could be used by indexers for indexing documents. The most appropriate terms selected while indexing a bibliographic record are known as descriptors or preferred terms. These terms (descriptors) are the basis for any thesaurus and become major part of controlled vocabulary. A descriptor becomes starting point and it will guide the indexer to choose appropriate related terms and narrow terms.

2.2.2 Non-preferred Terms/Non-Descriptor

A thesaurus should also bespeak some terms that cannot be used by indexers and questers when a term is associated with synonyms. Such synonymous terms which cannot be used or assigned to a bibliographic record as subject heading are known as non-descriptors. These terms are known as non-descriptor or non-preferred terms. A non-preferred term in a thesaurus provides link to a most preferred term or descriptor to be assigned to a record. In other words, a non-descriptor in a thesaurus guides the indexer to choose appropriate and most preferred term.

2.2.3 Related Terms

A standard thesaurus, in addition to providing pointer to the most preferred term in lieu of non-preferred terms, also provides list of terms which are related to a specific descriptor or a preferred term. These terms are called related terms or RTs.

2.2.4 Semantic Relations

According to Weinberg (1998), the thesaurus structure embodies rigorous semantic relationships and reflects the principle of post-coordination of terms. Weinberg (1998) also opines that rigorous semantic relationships allow a user to enter the thesaurus and to identify the appropriate search term(s). Thesauri contain three types of semantic relationship:

√ equivalence

√ hierarchy

√ association

BT (Broader Term), NT (Narrower Term), USE, UF (Used For) are some of the examples of semantic relations in a thesaurus.

2.2.5 Meaning of USE and Used For (UF) in a Thesaurus

A non-preferred term is normally linked to a corresponding preferred term by a USE reference. The corresponding reference in the opposite direction i.e. UF (õUsed Forö). For example,

Gender Discrimination

UF: Sex Discrimination

Preferred term is õGender Discriminationö and the corresponding non-preferred term is õSex Discriminationö.

2.2.6 Scope Notes

A scope note in a thesaurus guides an indexer to understand exact meaning of a descriptor and assists him/her in appropriate selection. More the number of scope notes better will be the quality of a thesaurus. Scope notes are very important as most of the indexers are not subject experts. A scope note also improves the indexing skill of a professional and augments quality of index.

2.3   How to Build a Thesaurus

A thesaurus which only lists all the preferred and non-preferred terms is known as enumerative thesaurus. The building processes in a thesaurus include: collecting terms, modifying terms, decision for descriptor or non-descriptor, establishing semantic relations and scope notes for defining a concept, etc,. An attempt has been made to describe the steps involved in construction of a thesaurus.

2.3.1 Collecting Terms

First and foremost step in construction of a thesaurus is collecting a set of terms. While some of these terms, thus collected become preferred terms, rest of these may become non-preferred. Before collecting sets of terms, one has to decide the sources from which such terms are identified. It could be existing thesauri, indexes, dictionaries, glossaries, etc., or it could be extracted from the textual metadata such as title, abstract, full text, etc., or it could be derived from the

discussion with a subject expert. Generally, a term in a thesaurus should include nouns or noun phrases and it should exclude proper nouns.

2.3.2 Modification of Terms as Per the Local Requirements

Majority of the terms collected for a thesaurus may be used as nouns or as noun phrases and adjectives. While building a thesaurus, it is extremely important to use such terms which are most sought by the end users at the time of retrieval. Such terms may vary from one country to another. For example, the descriptor *Reservation Policy* which is most accepted term in India is known as *Affirmative Action* in United States of America (USA). Similarly, there will be spelling variations across the countries, especially in USA and United Kingdom (UK). The term *Labour* will be spelt as *Labor* in USA. Therefore, while constructing a thesaurus it is also important to take into cognition the above mentioned and similar variations.

2.3.3 Establishing Relations

The third step in building a thesaurus is establishing relationships between terms. There are three types of relationships in any thesaurus - equivalence, hierarchy and association. According to Aitchison et al., (2000), the equivalence relationship is generally established between a descriptor and a non-descriptor. She also opined that the hierarchical relationship deals with a topical term and subordinate terms to establish relationship between superior terms and hyponym term. This type of concordant relation is established using BTs (Broad Terms) and NTs (Narrow Terms) which in turn establish hierarchical relation in a thesaurus.

According to Weinberg (1998) associative relationship means two terms overlapping with each other with same meaning. The associative may be symmetrical, e.g., gold is related to money and money is related to gold asymmetrical, e.g., population control is related to family planning, but there is no related-term reference in the opposite direction. (Someone searching for family planning is unlikely to be interested in population control.)

2.3.4 Thesaurus Display Format

The final step in constructing a thesaurus is to create display format according to the nature of thesaurus. The thesaurus could be displayed in two ways I) Alphabetic Sequence and II) Classified Sequence. In alphabetical thesaurus, terms are arranged in one alphabetic order and its associated terms are displayed under each descriptor. In case of a classified thesaurus, all the terms related to same concept will appear at same place and entire thesaurus is arranged in faceted manner with all the relations. According to Soergel (1974), the relationships in an alphabetic thesaurus should be in a specific sequence. As per his recommendations, a descriptor need to appear first in the sequence followed by Scope Note (SN), Broad Term (BT), Narrow Term (NT), Related Term (RT) in the same consistent order.

2.4   Role of Thesaurus in Indexing

Generally, an index represents a concept elaborated in a piece of information. Indexing is done to describe or identify the document using preferred term of subject content. Thesaurus (Controlled vocabulary), classification system and subject

headings are three most recognized and widely accepted indexing tools. The major role of a thesaurus is to help indexers to understand general comprehension of the subject area, outline inter-relationships between concepts, and provide definitions of terms as described by Aitchison, Gilchrist & Bawden (2000). Use of thesaurus in indexing of specific collection improves the quality of information retrieval in a particular subject domain.

2.5 Conclusion

Thesaurus plays a vital role in indexing and its subsequent retrieval of indexed documents. A bibliographic database, print or electronic, contains large number of records which need to have surrogates for effective and exhaustive retrieval. A descriptor is a surrogate for subject which embodies thought content of a document. Therefore, the technique of indexing using a standard thesaurus with appropriate search terms would fetch highly relevant documents at the time of search. Thesaurus construction involves multiple tasks which need to be executed in a logical sequence. All the above mentioned steps have been followed while constructing the thesaurus for Indian social science literature under the present research study.

References

1. Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: a practical manual* (4. ed). London: Aslib.

2. Cleveland, D. B., & Cleveland, A. D. (2001). *Introduction to indexing and abstracting* (3rd ed). Englewood, Colo: Libraries Unlimited.

3. Craven, T. (2008). Thesaurus Construction. Retrieved June 13, 2016, from http://publish.uwo.ca/~craven/677/thesaur/main00.htm

4. ERIC - Education Resources Information Center. (2016). Retrieved June 13, 2016, from https://eric.ed.gov/

5. Redmond-Neal, A., Hlava, M. M. K., Milstead, J. L., & American Society for Information Science and Technology (Eds.). (2005). *ASIS&T thesaurus of information science, technology, and librarianship* (3rd ed). Medford, N.J: Published for the American Society for Information Science and Technology by Information Today, Inc.

6. Soergel, D. (1974). *Indexing languages and thesauri: Construction and maintenance*. Los Angeles: Melville Publishing Company.

7. Weinberg, B. H. (1998). *Thesaurus Design for Information Systems*. Seminar presented at the Vocabulary Links:// Thesaurus Design for Information Systems, New York. Retrieved from http://www.allegrotechindexing.com/article02.htm