

DESIGNING A MICRO-THESAURUS ON ISLAMIC STUDIES

Syed Mudsair Ahmed Shah

Library Assistant at Beerwa College

E-mail: smudasir7775@gmail.com

Tahir Ahmed Batt

Instructor / Library Assistant

SKUAST-K, Shalimar-191201

E-mail: tahirbatt2015@gmail.com

Abstract: Thesaurus is the result of the active attempts made to overcome the problems involved in using uncontrolled vocabularies in indexing and searching of information. It is a vocabulary control tool designed to ensure effective representation and effective retrieval of information. Multi-lingual, multi-cultural thesauri are gaining importance as tools for knowledge organization as people with different cultural and linguistic background continue to publish and seek information. Islamic literature is published in many languages and there is need to develop an multilingual thesaurus on Islamic studies which covers all those major languages in which Islamic literature is published so that a standard set of Islamic terms is made available for the storage and retrieval of Islamic literature .

Keywords: Thesaurus, Islamic studies, Religion and Islam

1.0 Introduction : The effective retrieval of information by an information retrieval system demands a control over the vocabulary, it uses for indexing and the vocabulary used by information seekers in formulating search queries. This is because the richness of natural language in word forms and variety gives rise to many problems such as problems posed by synonyms, homonyms, spelling variations, scattering of related concepts etc which ultimately result in the poor performance of any information storage and retrieval system. In order to overcome these problems many forms of controlled vocabularies have been developed such as classification schemes, list of subject headings and thesaurus. These are all vocabulary control devices with some differing characteristics. The classification schemes and list of subject headings are generally related to pre-coordination systems where as thesaurus is meant for post-coordinate systems.

The word thesaurus has been originated from the Greek language and symbolizes the concept of treasury or store house of knowledge. The UNESCO has defined the thesaurus as “a vocabulary of controlled indexing language formally organized so that the apriori relationships are made explicit” (as cited in Kumbhar, 2004). This definition lays emphasis on apriori relationships which are document independent. Thus a thesaurus may be defined as a list of terms which represent only concept not subject matter or it is a compilation of all isolate ideas that occur within a subject or group of subjects and represented by appropriate isolate terms arranged in alphabetical order.

There are thesaurus in many fields such as thesaurus of library and information science terms in the field of library and information science. Islamic studies is one such field which has yet to

produce a thesaurus. There are glossaries and other compilations which list Islamic terms but not at par with a thesaurus.

Islamic studies is the academic discipline which focuses on each and every issue of Islam and Muslims as such it has a wide scope. According to **Nanji (1997)** the field of Islamic studies will continue to be more diverse and encompassing in its scope than in the past... the Islamic literature is written in almost all major languages of the world. However, the basic terms representing various concepts are almost same owing their origin to Arabic language. Also it is preferred to transliterate Islamic terms rather translate them into other language. It is thus possible to bring such terms together in one script with a sort of standardization and bring related concepts and terms at one place to ensure effective storage and retrieval of Islamic literature. The designing of micro thesaurus on Islamic studies is an endeavor in this direction with a limited scope.

2.0 Statement of Problem

There is hardly any tool available which provides standard Islamic terms for indexing Islamic literature. The lack of standard vocabulary affects the storage and retrieval of Islamic literature badly. Thus the problem is focused to design a micro thesaurus on Islamic studies. This may help in developing a standard vocabulary for Islamic studies.

3.0 Scope

Islamic studies being a vast discipline focusing on each and every issue of Islam and Muslims. The present micro thesaurus on Islamic studies covers only five fundamentals of Islam which may be broadly categorized into two categories i.e. aqaid or beliefs and ibadah or worship.

4.0 Objective

The objective of the work is to design a micro thesaurus on Islamic studies.

5.0 Methodology

The design of micro thesaurus is based on deductive method. The terms were collected from various sources such as books, Encyclopedia, glossaries and from web resources, recorded in roman script using generally adopted transliteration standard. Later the terms were studied in wider context to establish relationships between them and to choose preferred terms. In this regard help has been taken from the subject expert also. Finally the structure of the thesaurus was worked out according to the rules put forth for the thesaurus construction.

6.0 Literature Review

The literature review regarding thesaurus and its various facets shows that various aspects of thesaurus such as construction usability and development have been studied, however almost no work has been done specific to construction of thesaurus on Islamic studies.

Blagden (1968) has conducted a research focusing on utilization of faceted techniques to construct the thesaurus of management terms.

Aithcison (1970) has given a description of 'thesaurofacet' which is a faceted classification and thesaurus covering engineering and related scientific technical and management subjects

Roberts(1984) discusses at length the theoretical and practical contributions which shaped the thesaurus concept for operational use and highlights the problems of thesaurus use.

Lpoz huertas (1997) demonstrates the need for thesauri to help users in their search for information in online information systems and suggest that design of automatic thesaurus structure must be based on what is relevant for users and generators of information within a specific subject domain.

Shiri and Revil (2000) reviews the recent developments relating to the application of thesauri in information organization and retrieval on world wide web. It describes some recent thesaurus projects undertaken to facilitate resource description and discovery and access on wide ranging information resources on the internet.

Jorna and devis (2001) stress the need for developing a new form of multilingual thesaurus to remove the language barrier. **Reti and steno (2004)** discusses standards, language and utility of the laurin thesaurus. It also discusses the categories of terms used in the thesaurus and possible applications of the thesaurus beyond searching and browsing.

Middleton (2004) has compared various thesauri to know the status of information professionals as a descriptor in various thesauri. The study demonstrates that terminology that describe the discipline is inadequately defined and there is an imprecise and diversified choice of descriptors in different languages.

Mcculocry(2005) discusses term selection and other issues of thesaurus construction. The paper suggests that a team of individuals must be established to advise on term selection and structural issues. Shiri and

Revie(2005) discusses the ways in which end users perceive a thesaurus enhanced search interface. The study reveals that academic staff viewed the function of a thesaurus as being useful for narrowing down a search and providing alternative search terms, while post graduates stressed the role of the thesaurus for broadening searches and providing new terms.

7.0 Introduction to Thesaurus

Thesaurus is the result of the active attempts made to overcome the problems involved in using uncontrolled vocabularies in indexing and searching of information. It is a vocabulary control tool designed to ensure effective representation and effective retrieval of information. An abundant vocabulary allows a depth of expression that gives humans the ability to communicate at a high level, but it also has a semantic complexity that opens the door to faulty understanding (**Cleveland and Cleveland, 2005**). The indexer as well as searcher in the process of information storage and retrieval always come across variety of terms representing the same concept, because of synonymous terms, spelling variants, common names, usage variants, acronyms etc. use of such words without a control on part of indexer and searcher leads to wasted effort and inadequate retrieval. To exercise control over the vocabulary used in a particular information storage and retrieval system, many vocabulary control devices have been developed, thesaurus is one such device with distinguishing features found in it. Thesaurus controls the vocabulary in variety of ways , it determines the specifics of language and bring related concepts together in a systematic fashion.

Definitions:

According to Aitchison, Gilchrist and Bardon (as cited in Kumbhar, 2004 a), a thesaurus is a vocabulary of controlled indexing language, formally organized so that a priori relationships between concepts are made explicit to be used in information retrieval systems ranging from card catalogue to the internet.

Some other definitions include

1. A thesaurus is an alphabetical list of terms and concepts representing the specialized vocabulary of a particular field (Ali, 2006)
2. A thesaurus is a listing of terms for indexing (which have not been pre-coordinated), showing relationships among terms, specifically the generic relationship of broader and narrower terms and other relationships of synonyms and related terms (Corea, Ofuendo, & Farooqi, eds. 1993).

A thesaurus differs from the subject headings list as a thesaurus contains the terms which represent only concepts whereas subject headings list is a complete listing of subjects, where a heading may have many terms because the subject it represents is known or referred to like that. Similarly the thesaurus differs from classification scheme in that the classification schemes are hierarchical with secondary alphabetical index whereas thesauri are generally alphabetical with hierarchical structure built in by the use of references.

The word thesaurus has been used for centuries to mean lexicon or treasury of words. The present usage of thesaurus goes back to 1852, when Peter Mark Roget for the first time developed a language, thesaurus namely "thesaurus of English words and phrases". In 1878, W. F. Pole thought to construct a vocabulary control device in information retrieval on the lines of Roget's thesaurus and in 1959 the first thesaurus used in information retrieval was developed at Du Pont in USA.

From 1960 onward there were fast developments in the concept of thesaurus. During this period the thesaurus was widely adopted for vocabulary control. Initially it was used only in manual post-coordinate indexing systems then in mechanized ISRS and now it is used even in pre-coordinate indexing systems and of course in computerized ISRS (Kumbhar, 2004 b)

7.1 Categories of Thesauri

The broader categories of thesauri are 1. Language Thesauri and 2. Information Retrieval Thesaurus, however Thesauri may also be classified as 1. Monolingual Thesauri, 2. Multilingual Thesauri and 3. Interlingual Thesauri.

The language thesaurus is primarily a dictionary of synonyms, though structurally it may differ, to some extent, from ordinary word dictionary. Roget's Thesaurus is a classic example of language thesaurus which may be considered as a tool for writing in English. The information retrieval thesaurus collects terms of a subject or field and organizes in such a manner so as to bring out the relation between concepts, to ensure the accurate, clear and consistent indexing and retrieval.

Monolingual thesaurus is a thesaurus that can be used for indexing only in one language. Whereas multilingual thesaurus is used for indexing and searching in several languages. There are two types of multilingual thesaurus in one the descriptors and preferred terms are all in one language in second type there are different versions of indexing language. The interlingual thesaurus is constructed by collecting terms of various languages and designing a thesaurus without paying attention to the language to which each term belongs. In an interlingual thesaurus the preferred term may belong to any language which is used in the thesaurus.

7.2 Purpose of a Thesaurus

The main objectives of an information retrieval thesaurus are.

1. To permit the indexer of documents to index (i.e. describe) more fully the information contained in documents and
2. To permit searcher of information to phrase an inquiry appropriate to the scope and degree of his immediate interest.

According to Foskett (as cited in Kumbhar, 2004 c) the purpose of a thesaurus are as follows:

1. To provide a map of a given field of knowledge, indicating how concepts or ideas about concept are related to one another . this helps an indexer or a searcher to understand the structure of the field.
2. To provide a standard vocabulary for a given subject field which will ensure that indexers are consistent when they are making index entries to an information storage and retrieval system.
3. To provide a system of references between terms which will ensure that only one term from a set of synonyms is used for indexing one concept.
4. To provide a guide for users of the system so that they choose the correct term for a subject search. This stresses the importance of cross references. If an indexer uses more than one synonyms in the same index for example in case of Islamic Literature; Salah, Nimaz and salat, then documents are liable to be indexed haphazardly under all of these; a searcher who chooses one and finds documents indexed there, will assume that he has found the correct term and will stop his search without knowing that there are other useful documents indexed under the other synonyms .
5. To locate new concept in a scheme of relationships with existing concept in a way that makes sense to users of the system.
6. To provide classified hierarchies so that a search can be broadened or narrowed systematically.
7. To provide means by which the use of terms in a given subject field may be standardized.

Soergel (1974) lists few more uses of thesaurus in addition to its use within an ISAR system such as :

1. In collecting data, it serves as guideline during the data collection process whether for statistical purpose or for research.

2. In decision making where one needs a classification of alternatives and factors affecting the choice among alternatives.
3. A thesaurus can be used as a terminological standard in writing articles, instructions, standards etc.
4. A multilingual thesaurus can be used as a special multilingual dictionary in its field.

7.3 Structure

The internal relation of terms (concepts) to one another constitute the structure of thesaurus. The basic elements in a thesaurus are words, terms or phrases which form the entries. Descriptors are used by indexers to describe the subject content of materials and non-descriptors which have the same or nearly the same meaning as their equivalent descriptors but cannot be assigned to materials. Each entry consists of a pack of terms. The following examples describe the structure of present thesaurus.

1. Akhirah

UF uqba

Alam akhirah

Akhirat

BT alam

NT janah

Bahisht

Jahanm

Dawzakh

RT qiyamah

Barzakh

Mizan

To express various relationships following symbols have been used in the thesaurus.

SN scope note serves to avoid ambiguities.

USE, UF show equivalence relationship.

BT-broader term display hierarchical super ordinate relation

NT- Narrower term display hierarchical subordinate relation.

RT- related term, displays associative relations.

The relationships among the terms can be broadly classified into three categories as

- 7.3.1 Equivalence relationship.
- 7.3.2 Hierarchical relationship.
- 7.3.3 Associative relationship.

7.3.1 Equivalence relationship

They generally indicate preferred terms of descriptors and thus perform the necessary function of controlling index vocabulary. USE is used for preferred term and UF for non preferred term.

Example

Nimaz.

USE Salah

Salah

UF Nimaz

7.3.2 Hierarchical Relationship

This relationship shows level of super ordination and subordination . BT is used for super ordinate term and NT for subordinate term

Example

Hajj

BT ibadah

NT Tawaf

Rami

Sai

7.3.3 Associative Relationship

This relationship is found between terms which are closely related conceptually but not hierarchically and are not members of equivalence set. The relationship is indicated by RT.

Example

Ruqu

BT Salah

RT Sajdah

Qdah

Tashahud

8.0 Thesaurus Construction

The development of a thesaurus is a major project. Before starting such a project one should determine whether There is an existing thesaurus suitable to the needs of given ISAR system or requiring only slight modifications or there is no thesaurus at all. In former case evaluation of the existing thesaurus can be made through indexing and retrieval experiments. In case the results of evaluation are negative and if there exist no thesaurus, a new thesaurus has to be developed or an existing thesaurus has to be modified.

Lancaster (as cited in kumbhar, 2004, d) has identified four approaches to thesaurus construction

1. Convert an existing vocabulary e.g convert a list of subject headings into a thesaurus.
2. Extract the vocabulary from an existing, more general thesaurus i.e create a micro thesaurus .
3. Generate the vocabulary empirically on the basis of indexing a representative set of documents.
4. Collect terms together from diverse sources, including glossaries and other publications and form subject specialists.

The present thesaurus has been constructed using fourth approach which is also known as deductive method

8.1 Islam and Islamic Studies and Overview

Among the great religions of the world islam enjoys the distinction of bearing a significant name, a name that points to its very essence, islam is an Arabic word meaning “submission” the root meaning of the word islam is to enter into peace and a muslim is one who makes his peace with God and man (Ali, n p) islam is the name of third and last of the monotheistic faiths that arose in the middle east coming after Judaism and Christianity (Ayoub, 2006). Islam includes many religious practices , adherents are generally required to observe the five pillars of islam which are five duties that unite muslims into a community . Shia muslims subscribe to eight ritual practices which substantially overlap with the five pillars. In addition to the five pillars, Islamic law (shariah) has developed a tradition of ruling that touch on virtually all aspect of life and society.

Islam’s fundamental theological concept is tawhid the belief that there is only one God. The first of the five pillars, tawhid is expressed in the shahadah which declares that there is no God but Allah and that Mohammed (s. a. w) is Allah’s messenger. The other fundamentals include salah, zakah, sawm, hajj. Shia muslims consider three additional practices essential to the religion of islam which are jihad, amrbilmaruf and nahianilmunkar.

Islamic studies is an academic discipline which focuses on each and every issue of Islam and Muslims (tahir, 1991) Islamic studies by its nature is interdisciplinary and multidisciplinary, it assimilates fields generally considered to be secular in the west such as Islamic science and Islamic economics.

The subject areas covered under Islamic studies include;

- Islam as religion (theology)
- Islamic history
- Islamic culture
- Islamic architecture
- Islamic economics
- Islamic politics
- Islamic philosophy
- Islamic sociology
- Comparative study of religions etc.

Before 1980. Islamic studies was generally studied as part of oriental studies or Asian studies. However, in recent years it has assumed an independent status. According to Nangi (1997) there are many possibilities open for adding to its subject matter and methods, including the role of Islam as a cultural force of great diversity, the history of rural, agricultural and mountainous people of the Muslim world and new interactions among Muslims now living in the West, Islam and woman etc.

Islamic literature is published in many languages and there is a need to develop an interlingual thesaurus on Islamic studies which covers all those major languages in which Islamic literature is published so that a standard set of Islamic terms is made available for the storage and retrieval of Islamic literature. However, such work is challenging and requires team effort. The present thesaurus on Islamic studies is a microthesaurus and covers only a part of Islamic studies, i.e. theology. The terms belong to three languages: Arabic, Urdu, and Persian written in Roman script. There is scope to include English language also and construct a thesaurus which may help to index Islamic literature in any of these languages.

9.0 Conclusion

Multi-lingual, multi-cultural thesauri are gaining importance as tools for knowledge organization as people with different cultural and linguistic backgrounds continue to publish and seek information. There have been suggestions that presentation and structure of thesaurus should be language independent and efforts be made to adopt a semantic classificatory structure especially in a multilingual thesaurus. An important issue that has a bearing on the design of tools for knowledge organization including multilingual thesauri is that related to the uses to which such tools can be put in the emerging information environment. It has been suggested earlier that multilingual thesauri in culture-specific domains facilitate inter-cultural communication and comparative studies. In the context of the changing information landscape and growing importance of distributed digital resources on the Web, it is possible that tools for knowledge organization could be put to use in organizing, indexing and searching such resources. Traditionally the focus in research in the area of developing tools for Knowledge Organization such as thesauri and schemes of classification has been on the construction of such tools. In view of what has been discussed above it may be useful, while designing such tools, to focus on the use and possible applications of such tools. For example, the structure of a multilingual thesaurus may have a bearing on the design of search interfaces.

10.0 Reference

1. Ali, Muhammad. (n.p). The religion of Islamic discussion of the sources. *Principles and practices of Islam* (p.2) New Delhi: S Chand.
2. Atchison, Jea (1970). The thesauri facet: *A multiple retrieval language tool*. *Journal of Documentation* 26(2), 144-150. Retrieved April 21, 2008 from <http://www.emeraldinsight.com/10.1108/e-6026822>.
3. Ayoub, Mohammad, M. (2008), *Islam: Faith and History* (p.8). England: One world publishers.
4. Blagdon, J.F. (1968). Thesaurus compilation method: A literature review. *ASLIB Proceedings* 20(8), 345-359. Retrieved April 21, 2008 from <http://www.emeraldinsight.com/10.1108/002204/04/0534194>.

5. Corea, Ishavari, Ojavando, Gaddavid., & Faruqi, Khalid Kamal. (Eds) (1993). Encyclopedic of library and information science. v.10. (p.3115). New Delhi: Akshdeep Publishing.
6. Cleveland, Donald B. & Cleveland, Ana. D. (2005). Introduction to indexing and abstracting. (p.35) Colorado: Greenwood Publishing.
7. Kumbhar, Rajandra, (2004 a). Thesaurus of library and information science terms (p.26). New Delhi: Ess Ess Publications.
8. Kumbhar, Rajandra, (2004 b). Thesaurus of library and information science terms (p.27). New Delhi: Ess Ess Publications.
9. Kumbhar, Rajandra, (2004 c). Thesaurus of library and information science terms (p.34). New Delhi: Ess Ess Publications.
10. Nangi, Azim. (1997). Mapping Islamic studies: genealogy continuity and change. (p.5). Retrieved April 19, 2008 from [http://books.google.co.in/books?id=\]ojotwG87bxoc&pg=pa2198jpg=pa2198dq=islamic+studies+term&source](http://books.google.co.in/books?id=]ojotwG87bxoc&pg=pa2198jpg=pa2198dq=islamic+studies+term&source).
11. Shiri, Ali., & Crawford. (2005). Usability and user perception of a thesaurus-enhanced search interface, *Journal of Documentation* 61(5), 640-656. Retrieved April 21, 2008 from <http://www.emeraldinsight.com/10.1108/002204/04/0534194>.
12. Soergel, Dagabert. (1974) *Indexing language and Thesauri: Construction and maintenance*. (p.8). California: Melville publishing.
13. Taher, Mohammad. (1991). *Islamic studies in India: A survey of human, intuitional and documentary sources*. (p.7) New Delhi: Concept publishing.