

Knowledge Representation in the Spiritual and Cultural Domains

K. S. Raghavan*

DRTC, Indian Statistical Institute, Bangalore 560 059, India

Abstract: Knowledge organization systems such as thesauri are intended to structure and represent concepts in a domain in a manner that facilitates their effective employment in knowledge representation, i.e. representation of subjects of information resources (Classification and indexing) and users' queries (search and retrieval). Representation of 'aboutness' of a discourse and of the concepts in it is, to some extent, a function of the nature and characteristics of the 'domain'. Design of multilingual thesauri for multicultural domains such as 'spirituality' present many problems – both structural and semantic. Based on the experience in working on two multilingual thesauri in culture-specific domains, the principal issues that arise have been identified and illustrated with a few examples. The strategies that have been adopted to overcome the problems are also explained. A few general principles that could serve as guidelines in building multilingual thesauri especially in culture-specific domains have been suggested.

Keywords: Multilingual thesauri, knowledge organization, humanities, spiritual and cultural domain.

1. INTRODUCTION

Information management (IM) is not an end in itself; it is intended to support and ensure that some other activity – research, teaching, planning, decision-making, etc - is carried out in a more efficient and effective manner. An important requirement for efficient and effective IM – irrespective of the domain – is a mechanism for adequate representation of the 'aboutness' of discourses - information resources and user queries. Traditionally designers of information retrieval systems have assigned controlled vocabulary- based metadata to manage information resources to ensure effective and efficient retrieval of these. To a certain extent the issues related to representation of 'aboutness' of a resource are a function of the nature and characteristics of the 'domain'. This paper discusses some of the major structural and semantic issues in the design of a multilingual thesaurus in culture-specific domains.

2. THE SOCIAL IMPORTANCE

We live in a culturally diverse world. Cultural differences among communities have come to the fore in recent years. Misunderstandings, lack of communication among communities with different cultural backgrounds and practices have even resulted in violence, terrorism and a number of other related problems and issues. The events in the not too distant past in Bosnia and Serbia, the violence in Sri Lanka, the violent terrorist activities in different parts of the world have all basically sprouted from a lack of effective communication among the different communities and groups involved. The need for effective communication and understanding among members of different cultural and linguistic groups has never been greater than at present times. The differences in the

perceptions of the same set of events and happenings by a Middle East medium such as Al Jazeera and a western news channel such as the B.B.C. highlights the importance of and the need for understanding and appreciating cultural diversity. The importance of bridging the communication gap has been highlighted in such forums as the Atocha Workshop. [1] It has been said: "A tragedy of failure communication is the widening gap between the West and the Muslim world. In newswire dated, November 20, 2002; a prominent U.S. public relations executive, declared that both the Western democracies and the Muslim nations have "consistently failed to effectively communicate to one another the core values and realities of their respective ways of life," made a forceful plea to global leaders today "to do the right thing and avoid decisions that will place in harm's way millions of young people -- the flower of the future -- as we stand on the brink of war." Success depends on how well we communicate" [2]. Even in the spheres of trade and commerce, globalization has resulted in a situation in which, more than ever before, people encounter individuals from other cultures – people of different races, religions, linguistic groups, and nationalities. There is, often, a certain amount of anxiety surrounding unfamiliar cultures. Governments and inter-governmental or international organizations are looking at issues associated with and arising from globalization and related phenomenon and are concerned with containing and effectively managing the ill effects, if any, of these. Many intergovernmental and international organizations including UNESCO have emphasized the importance of promoting and facilitating effective cross-cultural communication as a means of developing and promoting better understanding among different communities of the world. While information and communication technologies have facilitated greater cross-cultural communication, they have also added to the growing global tensions especially when such technologies are used to whip up emotional issues centered on race, religion, culture and language. All these are highly emotional

*Address correspondence to this author at the DRTC, Indian Statistical Institute, Bangalore 560 059, India; Tel: +91-80-28483002/3006 Ext. 492; Fax: +91-80-28482711, +91-80-28484265; E-mails: raghavan@isibang.ac.in; ksragav@hotmail.com

issues and effective cross-cultural communication is seen as a way of reducing tension, mistrust and overcoming some other related problems that have increasingly come to affect peace and stability in many parts of the world. With the increasing pressures and opportunities resulting from globalization and the growing economic relations between nations, cross cultural communication is also a practical necessity. Cross-cultural communication should help communities and individuals understand what we as human beings have in common and also assist us in understanding the diversity which underlies our languages, life styles, practices and customs. That there is a spiritual dimension to human existence providing, for people of all faiths, the meaning and purpose of life needs to be understood and respected. Recognition of spiritual and religious needs of different cultural groups is an essential pre-requisite for this. As such in the emerging socio-political reality, accommodating spiritual pluralism, promoting exchange of ideas and creating opportunities for inter-faith dialogues will play an important role.

Spirituality is primarily concerned with matters of spirit – a concept that is widely associated with an individual’s efforts to attain the ultimate reality; and ‘*spiritual*’ is contrasted with the physical and material. It is an interdisciplinary area of Mysticism, Religion, Philosophy and a few other related areas. Spirituality often relates to the paranormal, the divine, etc. and as a domain it encompasses faiths, beliefs, religions, spiritualities, cults, cultures, races, meditation, mysticism, etc. There appears to be a growing volume of information on the subjects of cross-cultural, spirituality and multi-faith related studies. There is also interest in research into ‘*spiritual information*’ [3] and comparative and cross-cultural studies. A Google search for “*Multi faith studies*” retrieved over 24000 hits and one for “*spirituality and religion*”, nearly a quarter million! In recent years ‘*Intercultural relations*’ has even emerged as a relatively new formal sub-discipline dealing primarily with the ability of individuals to get along with others, especially those from a different cultural background, understanding social relationships and the way other cultures work. ‘*Intercultural relations*’ is again a fuzzy domain with inputs from many areas. The very nature of these domains is different from the more established natural and social sciences. Thus representing and organizing information in the domain presents many problems and challenges. In recent years a few projects have been initiated in India for designing and developing useful information products and information handling tools that facilitate and support comparative faith studies [4-6]. The objectives of this paper are to:

- Examine the issues relevant to building multilingual tools for knowledge organization in culture-specific domains;
- Explore the possibility of value addition to these tools using online lexical tools; and
- Suggest a few general guidelines for building multilingual thesauri

3. CONCEPTS IN SPIRITUAL AND CULTURAL DOMAINS:

“Concepts are the glue that holds our mental world together Our concepts embody much of our knowledge of

the world, telling us what things are and what properties they have” [7]. A concept is a mental representation that stands collectively for all the meaningful statements that could be made about an entity. Classification and indexing languages are artificial languages in which concepts are represented using symbols and / or descriptors. Information systems that employ thesauri or classification systems, in effect, attempt to codify reality into formal definitions. To that extent representations of reality are somewhat fixed until the scope of the descriptor itself is changed. Knowledge representations incorporated in these tools are based on assumptions regarding the practical need or purpose of the representation. Classifications and thesauri are largely limited to displaying only logical relations between concepts which are considered permanent features of the concepts so connected and are considered to be valid irrespective of the context (E.g. class inclusion or hierarchical relation, whole-part relation, etc). Even so they have an important role to play in enhancement of search in information retrieval. For example, the use of Library of Congress Subject Headings (LCSH) as a tool for interoperability among metadata schemes and for integrating multiple information repositories has been emphasized [8]. Yet, it is generally conceded that most users do not make full use of the power of such vocabulary control tools. Further, such general tools are quite inadequate for specialized areas such as spiritual and cultural domains and more specialized tools for knowledge representation are not widely available for these domains. Before we examine the issues related to building of multilingual thesauri for such domains, it is important to have an idea of the nature of concepts encountered in these areas.

The goal of spirituality as a way of life could be assumed to be to attain the ‘Absolute’ or the ‘Ultimate Reality’. It is largely concerned with the pursuit of communion with, identity with, or conscious awareness of an ultimate reality through direct experience, intuition, instinct or insight. Spirituality and Mysticism also focus on the attributes of the ‘ultimate reality’ and those of the ‘seeker’. There is also a substantial amount of literature on the processes by which, it is believed that, the union with the ‘absolute’ can be achieved. In recent years there has emerged a considerable amount of interest in comparative studies. Literature on Spiritual Mysticism in different religions and cultures is necessarily in different languages – primarily in the languages native to that particular religion. For example, much of Sufi mysticism literature is in Persian (and Arabic) just as much of Hindu mysticism literature is in Sanskrit (and other Indian languages). Also the concepts are somewhat unique to the context of the specific philosophy or religion such as Hinduism, Christianity, Islam, etc. While there are some common broad categories of concepts cutting across religions, a characteristic of the concepts in these domains that is particularly relevant to the task of developing multilingual thesauri for the domain is that a large number of concepts are those that have some meaning in the life of the members of the community belonging to the culture or religious group. This is understandable as the growth of concepts in the domains of Spirituality and Mysticism and the formation of lexemes in a language to denote the concepts are closely associated with the culture and lifestyle of the members of a particular religious or linguistic group. The result is that many of the concepts

Table 1. Categories of Concepts

Personality (P) Elements	Property (M) Elements	Action (E) Elements
Ultimate Reality (God, Allah, Paramatma, etc.) Deities Avatars Saints (By Religion) Alwars and Nayanmars <ul style="list-style-type: none"> • Alwars • Nayanmars Messiahs Seeker (Individual, Devotee) (By Religion) <ul style="list-style-type: none"> • Hindu • Christian • Muslim 	Attributes of the Ultimate Reality: <ul style="list-style-type: none"> • Uncreated (No beginning, No end) • All-pervading, Omnipresent • Unbounded, Infinite Attributes of the Seeker <ul style="list-style-type: none"> • Egoism • Selfishness • Possessiveness • Hatredfulness • Blinded by Maya • Paranormal behaviour <ul style="list-style-type: none"> ○ Occultism ○ Clairvoyance ○ Telepathy ○ Telekinesis 	Processes and Methods of attaining Reality <ul style="list-style-type: none"> • By Path <ul style="list-style-type: none"> ○ Devotion or Bhakti ○ Meditation

have an adequate and appropriate representation (in the form of suitable lexemes) often only in the language(s) native to a particular religion or philosophy. Also the concepts encountered in and associated with such culture-specific domains are abstract in nature and rarely can we relate these concepts to concrete referents.

Some of the more important broad categories of concepts encountered were:

- The ‘Reality’ and its attributes such as uncreated, all-pervading, omniscient, omnipotent, eternal, etc.,
- The attributes of the ideal seeker
- Connecting with the Reality
- Re-generation, transformation of the seeker, and
- Union with the Reality – Release and liberation

It was sought to classify and group these categories of concepts in terms of the framework of the Fundamental Categories enunciated by S. R. Ranganathan. The reason for this was that categorization provided the basic advantage of facilitating the mapping of the concepts into a hierarchical taxonomic structure. An idea of the nature of domain concepts that need to be represented in the different languages of the multilingual thesaurus can be obtained from Table 1 (Concept types are only illustrative and not exhaustive).

4. THE NEED FOR MULTILINGUAL THESAURI

In a comparative study of multilingual thesauri, InfoDEFT and Esser’s EXPO 2000 thesauri, Jorna and Davies remarked that: “multilingual tools are getting importance as increasingly diverse groups from different cultural and linguistic backgrounds seek access to equally diverse pieces of information” [9]. Two related projects triggered the need for the multilingual thesauri discussed in this paper. A multilingual multimedia information product, viz., the OM Information Service (OMIS) that ports selected extracts from the writings and sayings of spiritual leaders, saints, mystics and scholars to a CD (or to the Web) was developed with a view

to assist, among other things, comparative religious studies. The OMIS database responds to such queries as:

- a. Retrieving extracts from classics and commentaries related to similar concepts in different religions (E.g. mOksha in Hinduism and Salvation in Christianity)
- b. What are the different interpretations of the notion of ‘karma’?
- c. Sayings of Sikh Gurus and Sufi saints on the importance of ‘prayer’

The objective of OMIS was to serve as a “preliminary source of information for comparative and analytical studies, to prepare discourses and lectures, and to support and supplement attempts to understand the concepts and precepts of different faiths and cultures”[10]. More recently, the Government of India accorded classical language status to Tamil, Kannada and Telugu languages belonging to the Dravidian family of languages. The Tamil language was given this status in 2005 while the other two languages were added to the list of classical languages in 2008. Several projects have been initiated including the development of a *Digital Library of Tamil Classics*. This project is being seen as resulting in a major facility to help search and retrieve texts and passages from classical Tamil literature (of the Sangam period). Similar projects for the other languages are also likely to be initiated. A considerable proportion of this classical literature relates to such subjects as spirituality, mysticism and culture studies.

OMIS and the proposal to digitize classical literature in Tamil (and possibly in the other two languages in the foreseeable future) have also generated a considerable amount of interest in cross-cultural and comparative studies. In fact an important goal of these ongoing projects is to encourage ‘comparative studies’. It was realized that a multilingual thesaurus would facilitate not only indexing of and information retrieval from such digital resources, but will also assist comparative studies. This aspect had to be kept in mind in

the efforts to build the multilingual thesauri. Two thesauri – one for Spiritual Mysticism and one for Tamil Studies – are being developed. The thesauri are in different stages of development and the progress in respect of these has been reported elsewhere. However, it has been possible to identify some of the principal issues as also the characteristics of the domain that may impact design of information systems. Examples from these tools being developed will be used here to illustrate the issues encountered and the strategies adopted.

5. THE METHODOLOGY

Knowledge representation is constrained by contexts of language, culture and discipline. A piece of text in a particular language, except probably in science and technology, is not necessarily fully understood and comprehended by one other than a native speaker of the language. Meaning and understanding in Humanities (and to a lesser extent in the Social Sciences) are heavily dependent on understanding the context. There are three general approaches to building a multilingual thesaurus:

- Building a new thesaurus from the bottom up
 - Starting with one language and adding other language(s)
 - Starting with more than one language simultaneously
- Combining existing thesauri
 - Merging two or more existing thesauri in two or more different languages into a new multilingual thesaurus
 - Linking to each other two or more existing thesauri
- Translating an existing thesaurus in a language into one or more other languages.

The second and third approaches pre-suppose the existence of a thesaurus in the domain under consideration. In the areas of Spiritual Mysticism and Culture studies that were the main focus of the work reported here, since no thesaurus in any of the languages existed, the thesauri had to be built ‘*bottom-up*’. The candidate concepts (candidate terms) for the thesauri were gathered from a wide range of resources – both print and e-resources including several useful web-based resources. The major issues and strategies adopted are discussed in the following paragraphs.

6. DISCUSSION AND SUGGESTIONS

Quite early in the development of the thesauri the importance of defining a source language was realized. A general guideline that should be valid and applicable for any bilingual or multilingual thesaurus in a culture-specific domain is to start with the original language in which much of the domain literature appeared to begin with. This is helpful as most of the domain concepts will have appropriate lexemes in the source language. In the two thesauri referred to in this paper, the source languages were Persian (for Spiritual Mysticism) and Tamil (for Tamil Studies) respectively.

An important structural issue encountered was the one related to the structure of the thesaurus. Ideally a multilingual thesaurus should adopt a display providing parallel and symmetric displays of concepts, concept hierarchies and related concepts in all the languages of the thesaurus. How-

ever, this was found to be difficult to implement in spiritual and cultural domains. Symmetrical hierarchical displays presume the availability of equivalent concepts and terms for all the concepts or terms present in the source language hierarchy in the other languages of the thesaurus. In the spiritual and cultural domains many of the concepts are so unique to a particular culture, religion or philosophy, it was difficult to build parallel hierarchies with a one-to-one correspondence of concepts or terms in the different languages of the thesaurus. Symmetrical hierarchies of concepts may not exist. The strategy was to adopt a more pragmatic approach and to build separate hierarchies (as applicable to the domain and language) for the different languages. In other words the complete hierarchy for a descriptor indicating all conceptually related terms (synonyms, broader and narrower terms and related terms) was given only in the language of the descriptor. In the other languages of the multilingual thesaurus only the equivalent terms (quite often near equivalent terms as exact equivalence was difficult to find) were given which were hyperlinked to the corresponding records that provided full hierarchical displays of all semantically related terms in those languages. (see Fig. 1 in which EE indicates the equivalent English term and EV indicates the equivalent Vedic Sanskrit term for the descriptor in Persian or Arabic).

6.1. Semantic Issues

A thesaurus generally handles synonymy, hierarchical relations (Broader and narrower terms) and lateral relations. Two terms are considered equivalent if they represent the same concept. Intra-language synonymy is handled in a thesaurus by means of USE and UF links. Bilingual and multilingual thesauri also need to handle inter-language synonymy. A frequent problem in the spiritual and cultural domains is the non-availability of appropriate terms for concepts in languages other than the source language. In practice different kinds of strategies, depending on the appropriateness to the situation being handled, were employed:

- To employ the semantically nearest equivalent term (in the target language)
- To employ an appropriate phrase or compound term that represented the concept in the target language
- To employ the term in the source language itself; this was both necessary and useful in all cases of terms denoting concepts that have almost come to be accepted as proper nouns.

While there were many concepts in the spiritual and cultural domains common to many faiths and cultures, e.g., God, soul, prayer, after-life, renunciation, bliss, etc. the terms used in the different faiths and cultures are not exact equivalents and are not coterminous in what they represent. To give an example, in Sufi writings (Persian and Arabic) the term ‘QALANDAR’ denotes a dervish who does not recognize outward mystical form or convention. The nearest term in Sanskrit is ‘sanyasin’. The terms ‘Salvation’, ‘Moksha’, and ‘Nirvana’ (in Christianity, Hinduism, and Jainism and Buddhism respectively) denote closely similar concepts, but are not exact equivalents. Terms in the spiritual and cultural domains also have a large number of closely related terms (RTs) for many of which exact equivalents are not available in the other languages. The record shown in Fig.

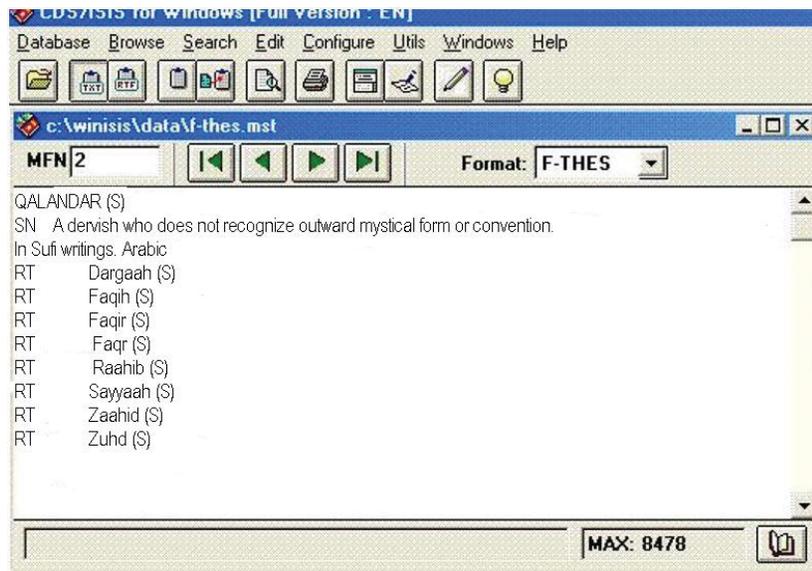


Fig. (1). A Sample Record.



Fig. (2). RTs in Culture-Specific Domains.

(1) for a Sufi term in the thesaurus is indicative of these issues.

Another major semantic issue that was frequently encountered was the presence, especially in Indian languages, of a large number of terms with closely similar meanings. This suggested the existence of a large number of RTs. This issue was particularly encountered in the bilingual Tamil-English thesaurus for Tamil Studies. The screenshot in Fig. (2) displays a relevant thesaurus record.

The partial record reproduced above is indicative of some of the complexities and difficulties involved in handling the large number RTs that characterize the spiritual and cultural domains. The web of relations in a thesaurus for any given concept (represented by a term) should ideally represent the collective world view about the concept. There are limits to the extent to which an information retrieval thesaurus could display the complete semantic map of any given concept. There are also limits to the extent of description (scope note)

that a thesaurus record can provide. In the present thesaurus the near impossibility of comprehensively displaying all RTs to a given descriptor was realized. These two issues triggered experimenting with the use of online lexical resources available on the Web to supplement and complement the network of RTs. Every thesaurus record was hyperlinked to the search page of the *University of Madras online Tamil lexicon*. This was seen as a value addition process and the reaction of the potential end users of the digital library to this facility clearly suggested that this does a great deal to enhance the semantics of a thesaurus term and in supplementing and complementing the relation network displayed by the thesaurus. The screenshots below give a general idea of the kind of value addition that could be made to a thesaurus by linking the terms to relevant online lexical resource (see Figs. 3, 4, and 5). The use of the University of Madras Tamil lexicon was found to be particularly helpful as it provided for display of relevant records in the 'keyword in context' mode also in addition to the regular mode.

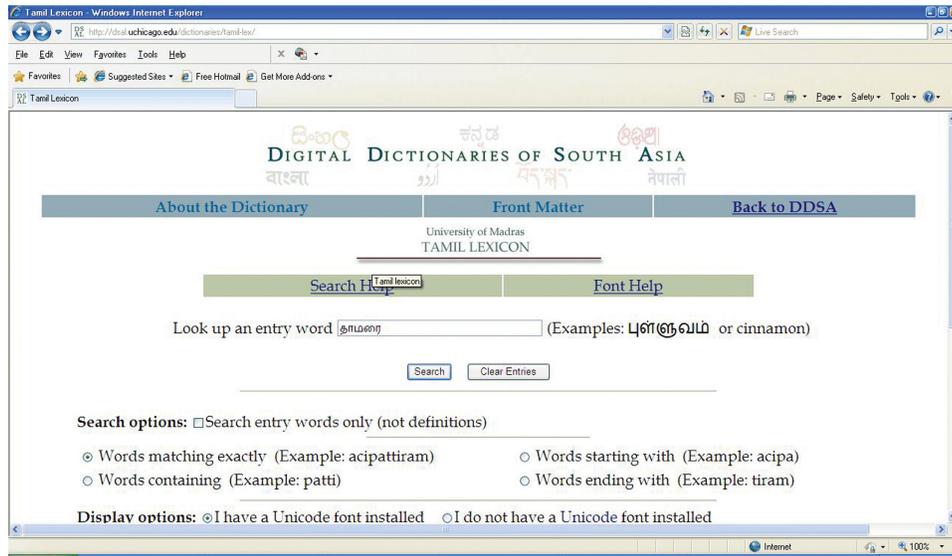


Fig. (3). Home Page of University of Madras Tamil Lexicon.

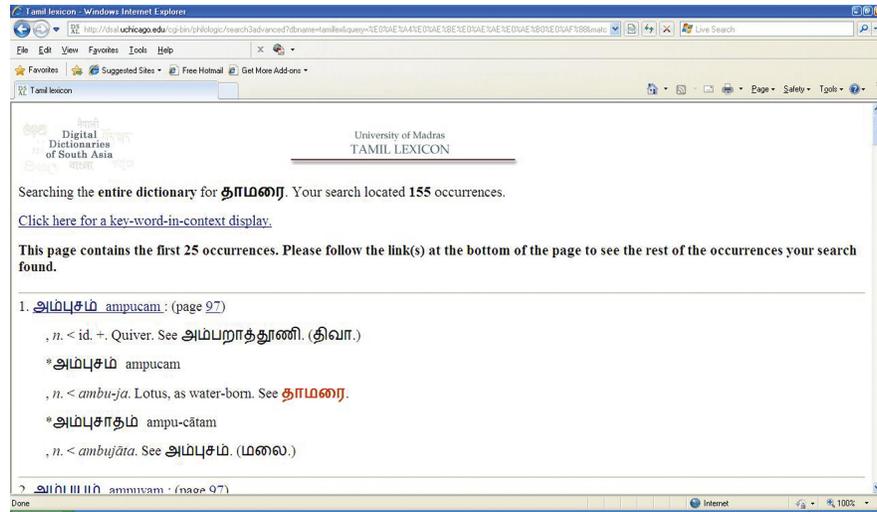


Fig. (4). Partial Display of Retrieved Lexicon Records.

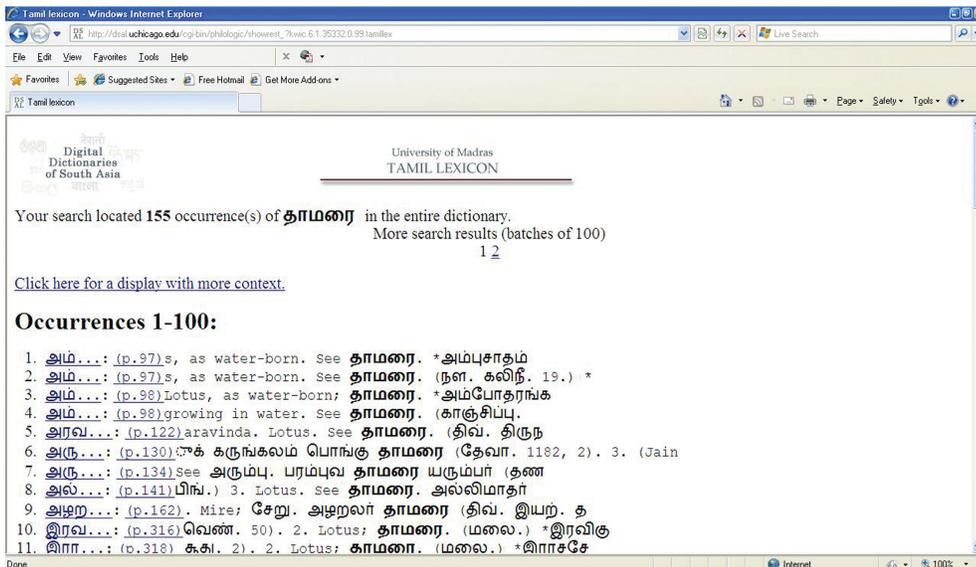


Fig. (5). Keyword-in-Context Mode Display.

The building of these thesauri also led to the identification and categorization of lateral relations (Associative relations). This has been reported elsewhere [11, 12].

CONCLUSIONS

Representation of concepts in the spiritual and cultural domains involves many issues. It is useful to adopt certain guidelines in the development of tools for knowledge organization such as multilingual thesauri in these areas. The use of available online lexical tools certainly adds value and enhances the semantics of the thesaurus records.

REFERENCES

- [1] http://ivythesis.typepad.com/term_paper_topics/2009/04/communication-gap.html (accessed Nov. 8, 2010).
- [2] <http://www.atochaworkshop.org/topics/from-raleigh-to-riyadbridging-the-communication-gap.html> (accessed Nov. 4, 2010).
- [3] Kari J. Informational uses of spiritual information: an analysis of messages reportedly transmitted by extra-physical means. *J Inform Sci* 2009; 35(4): 453-468
- [4] Raghavan K. S. Neelameghan A. A bilingual information retrieval thesaurus: Design and value addition with online lexical tools. In: Buchanan G, Masoodian M, Cunningham SJ. Eds. *Digital Libraries: Universal and Ubiquitous Access to Information: LNCS 5362*. Springer Verlag 2008; pp. 408-409.
- [5] Raghavan KS, Neelameghan A. Design and development of a bilingual thesaurus for classical tamil studies: Experiences and Issues. In: Arsenault C, Tennis J. Eds. *Culture and Identity in Knowledge Organization*: Ergon Verlag 2008. pp. 70-76.
- [6] Neelameghan A, Raghavan KS. An Online Multi-Lingual, Multi-Faith Thesaurus: A Progress Report on F-THES. *Webology* 2005; 2(4): Article # 19. Available from: <http://www.webology.ir/2005/v2n4/a19.html> (accessed Nov. 8, 2010).
- [7] Murphy GL. *The big book of concepts*: MIT Press 2002. p. 1.
- [8] Koch T, Day M. The role of classification schemes in Internet resource description and discovery. Available from: www.ukoln.ac.uk/metadata/desire/classification (accessed Nov. 8, 2010).
- [9] Jorna K, Davies, S. Multilingual thesauri for the modern world - no ideal solution? *J Document* 2001; 57(2): 284-295.
- [10] Neelameghan A. Lateral relations and links in multi-cultural, multi-media databases in the spiritual and religious domains: Some observations. *Inform Stud* 1998; 4 (4): 221-246.
- [11] Neelameghan A, Raghavan KS. Semantics of relationships in knowledge organization: lateral relationships. *SRELS J Inform Manage* 2005; 42 (4): 361-382.
- [12] Neelameghan A, Raghavan KS. Semantics of relationships in knowledge organization: lateral relationships: update 1. *Inform Stud* 2006; 12 (2): 115-122.

Received: March 30, 2010

Revised: May 07, 2010

Accepted: May 07, 2010

© K. S. Raghavan; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.