B.L. Vinod Kumar – Swami Vivekananda Yoga Anusandhana Samsthana, Bangalore, India
Khaiser Nikam – University of Mysore, Mysore, India

Sanskrit-English Bilingual Thesaurus for Yogic Sciences: A Case Study of Problems and Issues with Terms of Non-Latin Origin

Abstract
A machine-readable Sanskrit-English thesaurus for yogic and allied sciences is being developed using Greenstone Digital Library software (GSDL) at the S-VYASA deemed to-be university in Bangalore. This paper deals with the problems and issues that arose during construction of the bilingual thesaurus.

Introduction: Need for this Work
Swami Vivekananda Yoga Anusandhana Samsthana (S-VYASA) is an academic and research centre for yoga and allied sciences and has links with a number of institutions in India and abroad. Digital library of S-VYASA consists of an online catalogue of information resources – Books, theses, dissertations, reports, conference proceedings, and research papers. Full texts of research papers, theses and dissertations are accessible on their websites. Yoga is an ancient discipline and many of the early works are in Sanskrit e.g. Patanjali’s Yoga Sutras (PYS), Gheranda Samhita, Yoga Vasistha, Hatha Yoga pradipika, Shiva samhita, etc. Even today researchers publish in Sanskrit language on the subject. In recent times, however, there are books, reports and papers on yoga, its application and effects on human health in English and in some other languages of Latin origin. In this context the need for a vocabulary management tools was felt; hence the project for the design and development of Sanskrit – English bilingual thesaurus.

Review of Literature
A literature survey revealed that there are a number of papers on the design and development of digital information systems and knowledge organization tools (KOTs). To cite a few: Hudon (2001), Jorna and Davies (2001), Kwasnik and Rubin (2003), Landry (2004), Milstead (2001), and Molholt (2001) dealt with problems of multilingual thesaurus design and development; however most of them deal with terms of Latin origin. Neelameghan (2001), Neelameghan and Raghavan (2005), Raghavan and Neelameghan (2008) describe their work of preparing bilingual and multilingual thesauri, F-thes (Sanskrit, Farsi, English) and G-tamthes (Tamil and English). However in the humanities and yogic sciences in particular, there are only a few papers relating to terms of non-Latin origin (Neelameghan and Raghavan, 2005; Neelameghan, 2007).

Features of the Bilingual Thesaurus
The following requirements were kept in mind:
• It should be possible to search in Sanskrit and English;
• The software should be Unicode compliant and capable of handling Sanskrit (Devanagari) script;

• Enable hypertext linking within the thesaurus and with external databases and tools – dictionaries, lexicons, relevant descriptive texts online.
• It should provide for equivalent terms in both languages, if not near equivalent terms or transliterated terms.

Problems and Issues
For many Sanskrit / concept / terms exact English equivalent terms are not available; similarly, for the recent works about yoga in English, Sanskrit equivalent terms for some of the specialized English terms are not available. In such cases hypertext linking to the following alternatives can be used:

<table>
<thead>
<tr>
<th>Transliterated term(s)</th>
<th>Code e.g. Classification code</th>
<th>Scope note or descriptive definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoga terms in original texts available</td>
<td>not available</td>
<td>Use near equivalent terms; or Sanskrit transliterated in English</td>
</tr>
<tr>
<td>Yoga terms in modern research Equivalent Sanskrit terms not available</td>
<td>available</td>
<td>Use near equivalent terms; or English transliterated into Devanagari</td>
</tr>
</tbody>
</table>

User can search in Sanskrit or English; if a code is used one can give the code and can search all the languages in the thesaurus and external online lexicons and databases; codes can be a class numbers from a classification scheme.

Designing the Bilingual Thesaurus (Y-Thes)
The designing of the thesaurus involved the following tasks:
1. Study of the subject by examining the literature available on the subject, moving progressively from reference sources – dictionaries, glossaries, encyclopedias, etc. - to the more detailed material, such as, papers in periodicals, proceedings of seminars, conferences, symposia, etc.;
2. Creating a corpus of terms;
3. Identifying user needs from yoga library users profiles;
4. Selection of terms from the corpus in consultation with subject specialists and the users of information system;
5. Record each term in the format chosen as most convenient; The layout of the ultimate product is decided and the thesaurus is constructed accordingly.

Sources of Concepts / Terms
Corpus of terms - dictionaries, glossaries, lexicons - already exists. Relevant terms from the OM database, and the F-thes glossary were transferred. A collection of over 5000 books, peer-reviewed journals such as International Journal of Yoga (IJY) and International Journal of Yoga Therapy, theses and dissertations, research reports were used for identifying core concepts. Some select secondary sources were also used (see Annex 1).

Criteria for Selection of Software
The Greenstone Digital Library software (GSDL) for building the bilingual thesauri (Kumar and Nikam, 2011) was used as it was Unicode compliant and allowed hyperlinking within the thesaurus as well as to external sources.
### Database

**The fields and tags of the database**

<table>
<thead>
<tr>
<th>Tag</th>
<th>Field Name</th>
<th>Example of data entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>अष्टाङ्गयोगोऽ</td>
</tr>
<tr>
<td>2</td>
<td>EQS (Equivalent Sanskrit)</td>
<td>अष्टाङ्गयोग</td>
</tr>
<tr>
<td>3</td>
<td>EQS.USE</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>EQS.UF</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>EQS.BT</td>
<td>योग</td>
</tr>
<tr>
<td>6</td>
<td>EQS.NT</td>
<td>बहिरङ्ग साधना अन्तरङ्ग साधना अन्तरङ्ग साधना यम</td>
</tr>
<tr>
<td>7</td>
<td>EQS.RT</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>EQE (Equivalent English)</td>
<td>अष्टाङ्गयोग</td>
</tr>
<tr>
<td>9</td>
<td>Term</td>
<td>अष्टाङ्गयोग</td>
</tr>
<tr>
<td>10</td>
<td>ScopeNote</td>
<td>The Yogadarsana of Patanjali (200 B.C.) is named Astangayoga since it consists of eight limbs (asta = eight; anga = limb or steps). In Concise Encyclopaedia of Hinduism by Swami Harshananda. Non violence</td>
</tr>
<tr>
<td>11</td>
<td>USE</td>
<td>Ashtanga Yoga</td>
</tr>
<tr>
<td>12</td>
<td>UF</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>BT</td>
<td>BT: yoga</td>
</tr>
<tr>
<td>14</td>
<td>NT</td>
<td>NT: bahiraṅga sādhanā NT: antaraṅga sādhanā NT: antaraṭṭha sādhanā</td>
</tr>
<tr>
<td>15</td>
<td>RT</td>
<td>RT: Yama</td>
</tr>
</tbody>
</table>

**Example of Display**

Sanskrit term can be accessed by pressing the ‘Sanskrit term’ button in the navigation bar. This displays a list of terms in alphabetical sequence as shown in Fig. 1. Clicking on any of the displayed hyper link terms, say ‘यम’ will retrieve matching records in the pre-selected databases (See Fig. 2). English Descriptor(s) can be accessed by pressing the ‘English Descriptor’ button in the navigation bar. This displays a list of terms in alphabetical sequence as shown in Fig. 3.

**Fig. 1: Sanskrit terms (partial list) display by clicking Sanskrit term button**

![Image of Sanskrit terms display](image-url)
Utility and Value

In the Yoga Thesaurus search facility, users can type a Sanskrit term using on-screen keyboard. Search using on-screen Sanskrit keyboard, for keying in the search terms in languages other than English a Unicode compliant keyboard is necessary. The term(s) may be dragged from the list and dropped in the search box if necessary. The thesaurus is useful in domain mapping and for filling up the gaps in the vocabulary. In addition the possibility of using the thesaurus to develop a depth classification scheme for the domain is being explored.
Acknowledgment
We thank Ms. S.K. Lalitha of SRELS for help in the installation of GSDL.

References


Annex 1


