

Stuart J. Nelson

A Multilingual Vocabulary Project – Managing the Maintenance Environment

Forum Bibliotek Medycznych 2/1 (3), 449-453

2009

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

To obtain faculty participation in the submission process and to clarify its' own role, the Viikki Science Library has arranged workshops for and interviews with researchers on campus and instructed them personally. The library has also worked with the teachers to streamline the submission processes for students' theses. The contribution of faculty is absolutely necessary. Their attitudes and motivation is highly dependent on the rationality of the submission process, the support from the library and the usefulness of the repository in their daily work such as assessment of research, application for funds, assessment of students' theses etc.

The poster presents the role of the library to reach a better participation of faculty and students in DViikki by the cooperative means above and by technical means. Technical solutions undertaken are for instance loading metadata from other sources to DViikki, integration with the library catalogue and the catalogue for research output (Voyager) of the University of Helsinki and facilitation of digital archiving and preservation in the future through the National library of Finland.

MD Stuart J. Nelson

Bethesda – National Library of Medicine, Bethesda, Maryland USA

A MULTILINGUAL VOCABULARY PROJECT – MANAGING THE MAINTENANCE ENVIRONMENT

Abstract

The National Library of Medicine (NLM)'s MEDLINE/PubMed database includes over 14 million literature citations of articles written in 41 languages. International MEDLARS Centers, including those in Germany, Japan, Brazil, and France, as well as other national medical information centers have long produced translations of MeSH to make the vocabulary useful for non-English users. Various translations of Medical Subject Headings (MeSH) enable users not facile in English to identify articles that are of sufficient potential interest. Translations have generally been performed by individuals sufficiently well-versed in medical nomenclature in English and in the language to which they are translating.

A major concern of translators has been, and continues to be, the necessity of staying current with the annual editions of MeSH. To enable the translators earlier and more complete access to the development of MeSH, the MTMS was developed.

The Web-based interface of MTMS includes a variety of security measures to limit use authorized individuals. Privileges for translators are limited to insertion of terms in their own language, and to creation of new subordinate concepts. While the translator has the ability to browse MeSH descriptors, the translation interface has been designed for direct editing of concepts and terms only. The translator can quickly determine at a glance which MeSH terms are new, which still need to be translated, and which translated terms are waiting supervisor review and final approval. A special module of the interface

was designed for translators' supervisors, to enable them to review and authorize terms and concepts for each translator in their group.

The supervisor coordinates, reviews, and approves the work of the group for that language. After approval by the supervisor, staff at the National Library of Medicine makes a review of any added subordinate concepts to insure they are in the proper location. After that, all changes are approved to become an official part of the MeSH translation.

The National Library of Medicine (NLM)'s MEDLINE/PubMed is the premier international bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. MEDLINE/PubMed contains bibliographic citations and author abstracts from more than 5,000 biomedical journals published in 41 languages. The database contains over 14 million citations.

International MEDLARS Centers, including those in Germany, Japan, Brazil, and France, as well as other national medical information centers have long produced translations of Medical Subject Headings (MeSH) to make the vocabulary useful for non-English speaking users. Translations have generally been performed by individuals sufficiently well-versed in medical nomenclature in English and in the language to which they are translating. MeSH in translation can assist those not facile in English in identifying articles that seem to be of potential interest. A decision can then be weighed as to if the article warrants further effort to obtain a copy and or a translation of the article.

A major concern of translators has been, and continues to be, the necessity of staying current with the annual editions of MeSH. Each year, new descriptors are added to the MeSH vocabulary, existing descriptor class names are modified, and some descriptors are deleted. In addition, the entry or cross-reference terms are also subject to annual changes. The size of these descriptor changes is shown in the following table.

MeSH	Total descriptors	New descriptors	Changed descriptors	Deleted descriptors	Terms added	Terms moved to different descriptor	Total changes
2002	20,232	847	185	47	5,986	894	7,959
2003	21,079	1250	93	20	7,143	674	9,180
2004	22,329	666	109	20	5,719	519	7,032
2005	22,995	487	129	60	7,173	618	8,467
2006	23,885	933	188	43	14,342	511	16,017
2007	24,357	494	99	22	5,628	357	6,600

The MeSH Translation Maintenance System (MTMS) was developed to give the translators earlier and more complete access to MeSH developments in progress over the year-long cycle. Before the MTMS, translators received the complete MeSH vocabulary file in August. If they wished their translation included in the UMLS,

in order to meet processing UMLS deadlines, translators were asked to have all translations finished by November. In some years, more than 6,000 changes in MeSH must be translated. The MTMS provides the translator teams with ongoing access to the new MeSH version as it is being developed. The administrators can then assign work as time and staff are available and distribute the workflow in a more efficient manner.

MeSH translators have encountered difficulties with entry vocabulary as they maintain and update their translations to reflect changes in the annual version of MeSH. An entry term might move from one main heading to another main heading, or, more commonly, an entry term might become a new main heading. Translators are faced with difficulties in tracking these changes. Another problem arises for certain terms in other languages. There may be no exact English equivalent. In that case it may not be possible to identify the correct mapping to the MeSH descriptor or to concepts in other vocabulary databases, such as the UMLS Metathesaurus.

Underlying the structure of MeSH, NLM has developed and implemented a concept-centered vocabulary maintenance system for the vocabulary. These changes in the MeSH data structure support a multilingual Vocabulary. The underlying structure of MeSH changed effective with the 2000 version of the vocabulary. The new structure is centered on descriptors, concepts, and terms rather than only descriptors and terms. A descriptor is now defined as a class of concepts, and a concept as a class of synonymous terms.

A descriptor class consists of one or more concepts closely related to each other in meaning. For the purposes of indexing, retrieval, and organization of the literature, these concepts are best lumped together in one class. It has been recognized for some time that not every term that we might wish to explore is sufficiently distinct in meaning that it would serve well as a descriptor. For example, the *NISO standard for Monolingual Thesauri* talks of quasi-synonyms. These are terms that do not have the same meaning, such as “roughness” and “smoothness”, but are a means of addressing the same underlying phenomenon). Entry terms like “Isometric Exercise” are narrower in meaning than the main heading “Exercise”, but left in the exercise descriptor class because of the overlap in meaning with another entry term, “Aerobic Exercise.” The recognition of the nature of a descriptor as a class of concepts helps us to understand what we are dealing with.

Each descriptor has a preferred concept. The term that names that concept will be the preferred term of the preferred concept, and itself names the descriptor. Each subordinate concept also has a preferred term, as well as a specified relationship (broader, narrower, related) to the preferred concept. This structure allows such relationships to be expressed in a way that can be manipulated computationally. Furthermore, it allows each concept to carry its own unique attributes that have not

been previously represented. These include separate definitions, and translations into other languages.

While remaining invisible to the users of the system, the concept-centered structure supports a better understanding of the role of MeSH and of the composition of the thesaurus, and provides a useful method of representing the relationships between concepts. The MTMS extends this to create an interlingual database of translations. Each translated term is identified as a name of an existing concept, or as the name of a new concept created within the descriptor class. This database allows continual updating of the translations, as well as facilitating tracking of the changes within MeSH from one year to another.

In the MTMS, translated terms are provided as synonyms to existing concepts. For non-synonymous entry terms that are not present in the English version, but useful in the language of the translation, the translator creates a new concept. The concept belongs to a descriptor class, that of the main heading for which it was an appropriate entry term. In this case of a concept class for which there is no English synonym, a definition of the concept in English is required, so that translators using other languages can have the ability to include their terms in that concept class. In the case of creating a new subordinate concept, the required submission of a definition (in English) of the new concept supports both the translation of that term into other non-English languages, and enables proper maintenance when that descriptor class is edited by the MeSH staff.

The interface is Web-based to avoid the difficulties of trying to maintain multiple disparate clients and multiple versions of software. It also contains a variety of security measures to limit participation to authorized individuals.

Privileges for translators are limited to insertion of terms in their own language, and to creation of new subordinate concepts. While the translator has the ability to browse MeSH descriptors, the translation interface has been designed for direct editing of concepts and terms only. There are two different ways that the user can access the concepts and terms from the interface: (1) by searching the MeSH Tree Structures for descriptor names or (2) by searching for term names. For each method, there are two different language modes available: an English version, and a translated version that appears in each user's own language.

The interface uses color, boldface, and italic fonts in the display to convey the current status of the various descriptors, concepts, and terms. In this way the user can quickly determine at a glance which MeSH terms are new, which still need to be translated, and which translated terms are waiting supervisor review and final approval.

For each language to be incorporated into the MeSH maintenance environment, there is a team of translators and a supervisor. The language supervisor coordinates, reviews, and authorizes (releases) the work of that group of translators for that

language. A special module of the interface was designed for translators' supervisors, to enable them to review and authorize terms and concepts for each translator in their group. Once the supervisor authorizes the work, a member of the MeSH Section at the National Library of Medicine conducts a review of the proposed concepts that were created, to be sure they have been correctly placed in the correct descriptor class and are appropriately placed they, before they are approved to become an official part of MeSH.

The translation database requires the agreement and cooperation of the translators. As desired, previous translations can be loaded into the MTMS database from the UMLS Metathesaurus. Translations that have not been previously included in the UMLS Metathesaurus usually provide a term by term translation, which is then loaded into the MTMS. After the translations are loaded, translators would then be able to review areas in which the mapping from one term to another might be problematic, and to find the descriptors in MeSH for which there was no translated term. The display of translated terms in the concept structure allows finer shades of meaning to be fully represented.

The NLM provides the base vocabulary which is MeSH. The NLM also provides and maintains the client-server software, archives the translation data and provides subsets to each translating partner as specified by the translating institution. While there have been some experiments with building interfaces in languages other than English, none has been officially instituted.

Summary

When searching for information about a potentially relevant topic, it is often easier to search using the language with which one has the most facility. Translations of MeSH are valuable to persons not facile in English. The creation of the MeSH Translation Maintenance System enables correct mappings from one language to another to be maintained and enables translators to stay current with MeSH as it continues to be enhanced. The Web-based interface, closely managed maintenance environment, and adherence to modern standards, all provide a robust platform for an interlingual database of translations.