ISO 2788 + ISO 5964 + Much Energy = ISO 25964

by Stella G. Dextre Clarke

Information Standards

The most recent editions of ISO 2788 and ISO 5964, the international standards for monolingual and multilingual thesauri respectively, are dated 1986 and 1985. When one considers how much the information management environment has changed in the last two decades, this datedness is something of a disgrace to the information profession. Maybe the principles of how to build, maintain and use a thesaurus have retained their validity, but the context, technology and scale of application have altered so drastically that an update has to be long overdue.

International standards are reviewed every five years. A notice is sent out to all the members participating in the committee responsible for the standard. They vote on whether to revise, withdraw or simply confirm the standard for another five years. In the case of these particular standards the responsible committee is called *ISO TC 46/SC 9, Identification and Description*, which has 26 participating member countries, including Australia, France, Germany, United Kingdom, United States and too many others to name. When the standards came up for review around the turn of the century, all kept a low profile except for the British. And even they were very cautious in their approach to tackling the disgrace.

In 2000 a small band of thesaurus enthusiasts comprising the British Standards Institution (BSI) committee IDT/2/2/1 felt the shame was too much to bear, but quailed before the formidable task of establishing an international project. A more subtle approach was available. Since ISO 2788 and ISO 5964 are identical to the British Standards BS 5723 and BS 6723, they resolved to revise the British Standards first, then offer up the products to the international community. Seven years of hard slog followed, resulting

Stella G. Dextre Clarke is project leader and convenor, ISO NP 25964. She can be reached by email at stella<at>lukehouse.org

in the publication in 2008 of the last part of a spanking new standard known as BS 8723. It is spread across five parts as follows:

BS 8723: Structured vocabularies for information retrieval – Guide

Part 1: Definitions, symbols and abbreviations

Part 2: Thesauri

Part 3: Vocabularies other than thesauri

Part 4: Interoperability between vocabularies

Part 5: Exchange formats and protocols for interoperability

Of the above:

- Part 2 broadly covers all that was in BS 5723/ISO 2788 (plus a lot more);
- Part 4 covers all the multilingual content of BS 6723/ISO 5964, plus guidance on mapping between different types of vocabularies;
- Part 3 is completely new, describing classification schemes, subject heading lists, taxonomies, ontologies and name authority lists to help people understand the similarities and differences between them;
- Part 5 is new, too, providing a data model and format to facilitate data exchange.

When ISO 2788 and ISO 5964 came up for review again in early 2007, BS 8723 was duly offered up for international consideration, and ISO TC 46/SC9 voted to use it as the basis for revising the existing standards. And so project ISO NP 25964 was born.

Scope of ISO 25964

As stated in the project terms of reference, the proposal is to update and extend the scope of ISO 2788 and ISO 5964 to reflect the information environment of the 21st century. All of their existing scope will be retained and revised, and the following additional subjects will be added:

- Guidance on electronic functions and displays
- Functional specification for software to manage thesauri
- Guidelines for certain additional types of vocabulary (e.g. classification schemes, taxonomies)
- Interoperability (mapping) between vocabularies
- Formats for exchange of thesaurus data

This statement reflects what has already been realized in BS 8723, picking out only the most obvious changes. A more precise statement of the scope is under development as work proceeds.

Constitution and Procedure

A working group called WG8 (Structured Vocabularies) has been established to undertake the project. The standards bodies of 11 countries are currently represented on it: Canada, Denmark, France, Germany, Finland, New Zealand, South Africa, Sweden, United Kingdom, Ukraine and United States. The project leader and convenor is Stella Dextre Clarke of the United Kingdom. The secretariat for ISO TC46/SC9 and all its projects is the National Information Standards Organization (NISO) of the United States. NISO has greatly facilitated the work of WG8 by enabling regular teleconferences, as well as providing web space for sharing documents and discussion.

Progress to Date

WG8's first action was to review the drafts comprising the existing parts of BS8723 and collect the proposals of all the members for amending and enhancing them. These proposals were considered at an initial meeting held in Stockholm in May 2008. The most obvious change agreed was to rearrange the content in such a way that the five parts of BS8723 will be presented as two parts in ISO 25964, as follows:

ISO 25964. Thesauri and interoperability with other vocabularies

    Part 1: Thesauri for information retrieval
    Part 2: Interoperability with other vocabularies

The rationale for this change is as follows. Although Part 3 of BS 8723 covers vocabularies other than thesauri, it treats them in much less depth than it does thesauri. The main motivation for including them had been to

allow comparison of their key features and hence support interoperability when access to information resources requires use of multiple vocabulary types. For ISO 25964 it was felt that the unevenness of coverage should be reflected in the title and that all the material applying solely to thesauri should be in one document, whereas all the description of vocabularies other than thesauri should be put together in another document with advice on mapping between vocabularies and other aspects of interoperability.

Hence all the guidance on construction of thesauri, whether monolingual or multilingual, will be collected in ISO 25964 Part 1. The split between ISO 5964 and ISO 2788 had always involved a bit of repetition between the two, which had added to the total purchase price. The new arrangement will hopefully overcome both these problems. Not only that, but the document will include much new material relevant to such issues as the functions of thesauri in electronic environments, data modelling and formats and protocols for data exchange.

Following the Stockholm meeting, a rearranged draft of Part 1 has been prepared but still needs a lot more work. Writing a standard is much harder than writing a book, because it must represent a consensus among the experts and user communities involved. Two mini-groups, on multilingual issues and data modelling respectively, are currently thrashing out the more demanding challenges. October 30, 2008, is the target date for a revised draft to be circulated to the parent committee ISO TC 46/SC9. The process should lead on to publication of a Draft International Standard by April 30, 2009 – at which point feedback from anyone, anywhere, will be warmly invited.

Not Forgetting Part 2

At the time of this writing the energies of the working groups are focused on Part 1, but come November 2008, attention will turn to the knotty problems associated with use of more than one vocabulary type in networked systems. Mapping between vocabularies is rarely easy, but is very much needed in a society that increasingly expects to be able to access the whole world's resources, right now and in whatever language is chosen. ISO 25964-2 aspires to provide guidance in this difficult area, building on the foundations set out in BS 8723-4. A timetable will be established before the end of 2008.

But Where Does Z39.19 Fit In?

An American audience may be concerned about the relationship of the new proposal to its national standard for monolingual thesauri, ANSI/NISO Z39.19. The first edition of Z39.19 came out in 1974, the same year as the first edition of ISO 2788. Over the years, Z39.19 has been updated more regularly than ISO 2788, but the two have remained broadly aligned and compatible with each other. Work toward the latest (2005) edition of Z39.19 began at around the same time as the work on BS 8723, and the drafting committees in NISO and BSI regularly shared their documents, with the aim of avoiding conflicts. ANSI and NISO are represented on WG8 by Marcia Zeng of Kent State University, who very actively contributes comments from NISO committee members. So there is every prospect that the emerging international standard will complement and extend the content of Z39.19, especially in the areas of multilingual coverage and data exchange.

Participation

Sometimes the workings of the national and international standards bodies can seem quite opaque, their drafts protected by confidentiality and copyright constraints, their eventual publications protected by high prices too. But in reality there is a way in for people interested enough to approach their national standards body and make contact with the responsible committee. WG8 welcomes comments from the user communities on any aspect of the emerging standard. Participation will become all the easier when the first draft international standard is published in 2009. ■

In Conclusion

The professional disgrace of neglected standards is well on the way to banishment. ISO 2788 and ISO 5964 will before long retire gracefully from the scene, and the information community will have a new, sharper set of tools with which to access networked information resources and prepare for the Semantic Web. ■

Resources for Further Reading

BS 8723-1:2005 *Structured vocabularies for information retrieval – Guide – Definitions, symbols and abbreviations.* (2005). London: British Standards Institution.

BS 8723-2:2005 S*tructured vocabularies for information retrieval – Guide – Thesauri.* (2005). London: British Standards Institution.

BS 8723-3:2007 *Structured vocabularies for information retrieval – Guide – Vocabularies other than thesauri.* (2007). London: British Standards Institution.

BS 8723-4:2007 S*tructured vocabularies for information retrieval – Guide – Interoperability between vocabularies.* (2007). London: British Standards Institution.

DD 8723-5:2008 *Structured vocabularies for information retrieval – Guide – Formats and protocols for data exchange.* (2008). London: British Standards Institution.

Dextre Clarke, S. G. (2007). Evolution towards ISO 25964: An international standard with guidelines for thesauri and other types of controlled vocabulary. *Information Wissenschaft & Praxis, 58*(8), 441-444.

ANSI/NISO Z39.19-2005 *Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies.* Bethesda, MD: NISO Press. Also available at www.niso.org/standards/index.html.

TOP OF ARTICLE   < PREVIOUS PAGE