ISO 25964: a standard in support of KOS interoperability

Stella G Dextre Clarke

Project Leader, ISO NP 25964

Abstract

ISO 25964: Information and documentation - Thesauri and interoperability with other vocabularies will be a two-part standard that updates and revises the existing international standards for monolingual and multilingual thesauri. And not just that: it should also help thesauri to play their full part in access to networked resources, particularly in a Semantic Web context. Part 1 (dealing with the development and establishment of thesauri) is on track for publication in 2011, while Part 2 (focusing on interoperability) is scheduled for completion in 2012. A large part of the content is intended to standardize mapping between one vocabulary and another.

Introducing ISO 25964

ISO 25964: Information and documentation - Thesauri and interoperability with other vocabularies will have the following two parts:

Part 1: Thesauri for information retrieval

Part 2: Interoperability with other vocabularies

The standard is intended to update, revise and replace both ISO 2788 and ISO 5964 (the current international standards for monolingual and multilingual thesauri respectively) as well as BS 8723. The key difference between ISO 25964 and its precursor standards is its focus on interoperability. In other words, it has moved from the world of stand-alone printed thesauri to a networked scenario in which the thesaurus needs to be fit for seamless functionality amid today's expanding opportunities.

The drafting of ISO 25964 is undertaken by a working group (WG8) with members from 15 countries, a chairman from the UK, and a Secretariat run by NISO in the USA. The membership of the drafting committee is as follows:

Ingrid Berg, Sweden Javier Garcia Marco, Spain

Anders Cato, Sweden Alan Gilchrist, UK

Sylvie Dalbin, France Desislava Gulenova, Bulgaria

Johan De Smedt, Belgium Michele Hudon, Canada

Alina Deniau, France Eeva Karki, Finland

Stella Dextre Clarke, UK Daniel Kless, Germany

Axel Ermert, Germany Traugott Koch, Germany

Richard Light, UK Esther Scheven, Germany

Zhu Lijun, China Cherie Shum, New Zealand

Jutta Lindenthal, Germany Douglas Tudhope, UK

Marianne Lykke, Denmark Bernard Vatant, France

Hester Marais, South Africa Sun Wei, China

Dominique Menillet, France Leonard Will, UK

Per Nonboe, Denmark Zhengqiang Zhang, China

Ok Nam Park, South Korea Marcia Zeng, USA

Part 1: Thesauri for information retrieval

Part 1 of ISO 25964 deals entirely with thesauri (monolingual and multilingual), and is expected to be published in 2011. It covers:

- thesaurus content and construction, mono- or multi-lingual;
- guidance on applying facet analysis to thesauri;
- guidance on managing thesaurus development and maintenance;
- functional requirements for software to manage thesauri;
- a data model and derived XML schema.

While the two are not identical, the ISO 25964 data model is largely compatible with that of the W3C standard SKOS (Simple Knowledge Organization Schemes). Use of SKOS will therefore enable a great many thesauri compliant with ISO 25964 to be published on the World Wide Web.

Part 2: Interoperability with other vocabularies

The types of vocabulary in ISO 25964-2 (as Part 2 is known) include classification schemes, file plans, taxonomies, subject heading schemes, ontologies, terminologies, name authority lists and synonym rings. ISO 25964-2 does not attempt to standardize these other vocabularies: it will only describe them sufficiently to explain how to support interoperability with thesauri, especially how to build and implement mappings between each of these and thesauri. Behind-the-scenes mapping from one vocabulary to another is the technique which should allow a search phrased in the terms of one thesaurus to be converted to other vocabularies, some of them in completely different natural languages, to retrieve resources from all over the networks.

The standard sets out the principles and many practical examples of mapping between one thesaurus and another. It also deals with mapping between thesauri and the other vocabulary types. As well as describing mapping types in detail, it discusses structural models for mapping across vocabularies, deals with the extra complications of handling pre-coordination among

vocabulary concepts, and has some advice on data management. Drafting is very actively under way, so at the time of the ISKO-UK 2011 Conference inputs from all interested parties can still have influence on the outcome. Later in 2011 it is hoped that a draft will be circulated within the ISO system (coordinated by ISO SC46/SC9) and will be made available for public comment.

More detail on mapping types

A 'mapping' is defined in the latest draft of ISO 25964-2 as a 'relationship between a concept in one vocabulary and one or more concepts in another'. Table 1 shows the three basic mapping types recommended:

Mapping type	Tag	Example	
Equivalence	EQ	laptop computers EQ notebook computers	
Hierarchical	NM	roads NM streets	
	BM	streets BM roads	
Associative	RM	journals RM magazines	

Table 1: Basic mapping types in ISO 25964-2, with examples

The second column in Table 1 indicates the tag that should be used in mapping statements. Each of these tags has a parallel in the tags traditionally used for internal relationships, as recommended in ISO 2788, ISO 5964 and ISO 25964-1. Thus NM, BM and RM, standing for 'Narrower Mapping', 'Broader Mapping' and 'Related Mapping' respectively, correspond to the traditional tags NT, BT and RT. The tag EQ stands for Equivalence, and it corresponds to the traditional USE/UF tags used between preferred and non-preferred terms in a monolingual thesaurus.

There is a subtle difference between EQ on the one hand and USE/UF on the other. EQ marks a mapping between equivalent *concepts* in different vocabularies; whereas USE and UF mark relationships between equivalent *terms* in the same vocabulary. When two or more terms in the same vocabulary are deemed equivalent, one of them is given 'preferred' status and the others are non-preferred, hence we use the tag USE to show an equivalence relationship in the direction from non-preferred to preferred, while UF marks the direction from preferred to non-preferred. In the context of a mapping between concepts in two vocabularies, however, both concepts have equal status and only one tag is applicable, in either direction.

Despite the simple brevity of Table 1, ISO25964-2 accompanies these basic mapping types with additional recommendations that allow greater semantic specificity, for optional use. According to the draft standard, different degrees of equivalence can be specified, the equivalence mapping can be simple or compound, and it is possible to differentiate between three subtypes of hierarchical mapping. The various options for differentiation are summarized in Table 2. The content of this table is drawn from the latest committee draft of ISO 25964-2, but it should be noted that this is still subject to change up until approval of the final version.

Mapping type	e Optional differentiation	Examples
Equivalence		laptop computers EQ notebook computers
	Exact	mad cow disease =EQ bovine spongiform encephalopathy
	Inexact	lawns ~EQ turf
	Compound	
	Intersecting	women executives EQ women + executives
	Cumulative	inland waterways EQ rivers canals
Hierarchical		roads NM streets; streets BM roads
	Generic	rats BMG rodents; rodents NMG rats
	Instantial	Paris BMI capital cities; capital cities NMI Paris
	Partitive	fingers BPI hands; hands NPI fingers
Associative		journals RM magazines

Table 2: Summary of mapping types in ISO 25964-2, and options for more differentiation

Conclusion

ISO 25964 will bring us a functional combination of old and new. It will build on the established principles and conventions of the existing standards, while adding to these a new capability for performance in the environment of networks in general and the Semantic Web in particular.

In Part 1, the principal manifestation of this new emphasis is the data model, with recommendations for publication on the Web. Part 2 appears completely new, but in practice the types of mapping recommended are not radically different from what has gone before. The standard allows scope for some useful semantic distinctions on mapping types, without raising barriers to implementation at a simple level. The project to reach agreement on all of this is well under way, with good prospects of the standard reaching approval in 2012.

References

British Standards Institution. *BS 8723: Structured vocabularies for information retrieval - Guide*. (published in five separate parts between 2005 and 2008) London: British Standards Institution.

International Organization for Standardization. *ISO 2788-1986: Documentation - Guidelines for the establishment and development of monolingual thesauri.* 2nd ed. Geneva: International Organization for Standardization, 1986.

International Organization for Standardization. *ISO 5964-1985*: *Documentation - Guidelines for the establishment and development of multilingual thesauri*. Geneva: International Organization for Standardization, 1985.

International Organization for Standardization. *ISO/FDIS 25964-1: Information and documentation - thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval.*Geneva: International Organization for Standardization; Final Draft circulated April 2011.

International Organization for Standardization. *ISO DIS 25964-2: Information and documentation - thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*. Geneva: International Organization for Standardization. Draft to be circulated for public comment late 2011.

Miles, A., & Bechhofer, S, Eds. (2009). *SKOS Simple Knowledge Organization System Reference*. [W3C Recommendation.] Retrieved from http://www.w3.org/TR/skos-reference.

Corresponding author:

Stella Dextre Clarke can be contacted at stella@lukehouse.org