Legal Information Management, 10 (2010), pp. 43–45
© The British and Irish Association of Law Librarians
doi:10.1017/S1472669610000265

Thesaurus Standards on a Converging Track

Abstract: Stella Dextre Clarke, our Guest Editor, provides an update on the latest developments nationally and internationally on producing standards for the creation of controlled vocabularies. After many years of inaction the advent of the worldwide web has stimulated the need for wholesale revision of existing standards and the introduction of new ones.
Keywords: standards; thesauri

Introduction

Have you ever wished it was easier to enter just one query and it would be executed simultaneously across all the relevant databases, repositories and collections, without need of reformulation? The availability of the internet, the worldwide web and countless other networks has created a demand for seamless access across multiple heterogeneous resources and stimulated a massive update of the national and international standards for thesauri and other types of vocabulary. This article has the latest news of progress with the standards.

Background

A huge number and variety of thesauri were implemented in the seventies and eighties of the last century and regarded as highly effective retrieval tools, each for its own database. But after the second edition of ISO 2788 (See Box for details) was issued in 1986, the international committee responsible for its maintenance became dormant for two decades. ISO 2788 for monolingual thesauri and ISO 5964, the companion standard for multilingual thesauri, were adopted as national standards in many countries. In the UK they were issued as BS 5723 and BS 6723 respectively. After that only the USA continued to update its separate standard ANSI/NISO Z39-19.

The new millennium has breathed new life into the standards, however, not just in the UK but internationally. The growth of networks and networking has brought new challenges. By 2000 the need for interoperability between collections, databases and systems indexed with different vocabularies had become so pressing that updating began in the UK and the USA.

In the UK it was decided to bring the monolingual and multilingual issues together in one standard, so that instead of BS 5723 and BS 6723 we would have just one BS 8723. But the committee soon discovered that they had far too much material to cram into one publication. BS 8723 eventually came out in five parts. As well as updating all that was in the earlier standards, it incorporates descriptions of classification schemes, subject heading schemes, taxonomies, ontologies and name authority lists. The idea is to explain the principles of mapping between one vocabulary type and another. It also provides a completely new data model and XML schema as the format for exchanging thesaurus data. The model and schema, (but unfortunately not the rest of the standard!) can be accessed free of charge at http://schemas.bs8723.org/.

In the USA parallel work led to the publication of a 2005 edition of Z39.19. As well as thesauri, this edition covers other types of controlled vocabulary such as lists, synonym rings and taxonomies. It does not cover multilingual vocabularies or mappings.

Current international developments

In the information business national isolationism is doomed – our standards need to be international. So as soon as Part 5 of BS 8723 whose official number is DD8723-5 was completed in 2008, a working group was formed to adopt it as an international standard. Officially named ISO TC46/SC9/WG8, the group has representatives from 15 countries: Bulgaria, Canada, China, Denmark, France, Germany, Finland, Korea, New Zealand, South Africa, Spain, Sweden, UK, Ukraine and USA. The aim is to revise and update ISO 2788 and ISO 5964, basing the work on BS 8723.

International adoption sounds misleadingly simple. In real life, things are more complicated. This is the opportunity to challenge the fundamentals and rework the standard so that it meets the great diversity of needs and perspectives from the USA through Europe to China and New Zealand.

https://doi.org/10.1017/S1472669610000265 Published online by Cambridge University Press

Status of the current thesaurus standards		
Type	**Number and title**	**Where we are at**
International	ISO 2788-1986: Guidelines for the establishment and development of monolingual thesauri	Principles still sound, but text dreadfully out-of-date
	ISO 5964-1985: Guidelines for the establishment and development of multilingual thesauri	Principles still sound, but text dreadfully out-of-date
	ISO 25964: Thesauri and interoperability with other vocabularies Part 1: Thesauri for information retrieval Part 2: Interoperability with other vocabularies	This standard will revise and update ISO 2788 and ISO 5964, as well as expanding their scope. It will be based on BS 8723, but be re-arranged into just two parts. Part 1 will be published in 2010
International Draft	ISO/DIS 25964-1 Thesauri and interoperability with other vocabularies. Thesauri for information retrieval	Issued Nov 2009 and available for public comment until end Feb 2010. Obtain it from BSI.
National (UK)	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	The five parts were published between 2005 and 2008. Part 2 basically covers monolingual thesauri, updating ISO 2788. Part 3 applies to classification schemes, taxonomies and other vocabulary types. Part 4 covers multilingual thesauri and mapping between vocabularies, updating ISO 5964 Part 5 is new.
National (USA)	ANSI/NISO Z39.19-2005: Guidelines for the construction, format and management of monolingual controlled vocabularies	Although recently updated, Z39.19 does not cover interoperability issues and so the USA is participating in ISO 25964 development.
Worldwide	IFLA Professional Report 115: Guidelines for multilingual thesauri	Published 2009. Aims to supplement ISO 5964
Worldwide	SKOS Simple Knowledge Organization System Reference	W3C Recommendation issued 18 August 2009

Accordingly, ISO 25964 has a new title and re-arranges the BS 8723 material into just two parts:

Thesauri and interoperability with other vocabularies
Part 1: Thesauri for information retrieval
*Part 2: Interoperability with other vocabul*aries

Part 1 covers:

- Thesaurus content and construction, mono- or multilingual.
- Guidance on applying facet analysis to thesauri
- Guidance on managing thesaurus development and maintenance
- A functional specification for software to manage thesauri
- A data model and derived XML schema

A draft numbered ISO/DIS 25964-1 was issued for public comment in November 2009. Until the end of February you can order it from BSI at http://shop.bsigroup.com/. It is listed under 09/30165649 DC as well as ISO 25964-1. The hard copy costs £36 (£18 to BSI members) but online it can be viewed free of charge at http://drafts.bsigroup.com/.

The types of vocabulary in Part 2 will include classification schemes, file plans, taxonomies, subject heading schemes, ontologies, topic maps, terminologies, name authority lists and synonym rings. ISO25964 will not be attempting to standardise these other vocabularies; it will only be describing them sufficiently to explain how to build and implement mappings between each of these and

https://doi.org/10.1017/S1472669610000265 Published online by Cambridge University Press

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 languages, to retrieve resources from all over the networks.

Convergence not conflict

Today's information networks can be impeded by frustrating barriers if different communities set up incompatible standards. Hence the collaboration in this project of so many countries, including China and the USA, is very reassuring. The working group is also taking note of guidelines for multilingual thesauri recently published by IFLA, the International Federation of Library Associations.

For the environment of the worldwide web, another key standards body is the W3C consortium. Their standards such as SKOS (Simple Knowledge Organization Schemes) and OWL (Web Ontology Language) address interoperability and are planned to enable the forthcoming Semantic Web. SKOS is especially relevant as a standard for publishing KOSs such as thesauri, classification schemes and subject heading schemes on the web. It aims to enable exploitation of resources classified and/or indexed with existing schemes.

The good news is that the architects of SKOS have consistently been taking account of the work on BS 8723. Even now the SKOS community is considering extensions for better alignment with ISO 25964-1. Members of the information profession have an instinctive predisposition towards sharing resources. The prospects are therefore good for converging standards to free up the flow of information across the web, across communities and across languages.

Legal Information Management, 10 (2010), pp. 45–47
© The British and Irish Association of Law Librarians
doi:10.1017/S1472669610000034

The Real CIG – Who are They?

Abstract: Andrew Coburn the 2009 Chairman of the CILIP Cataloguing and Indexing Group describes its activities, including continuing professional development, in the form of meetings and courses; work on standards and the opportunities it provides for networking.
Keywords: cataloguing; indexing; professional organisations.

Introduction

The original CIG founded in 1965, is CILIP's Cataloguing & Indexing Group. While we regret the recent passing of the City Information Group, the Cataloguing & Indexing Group is now the only owner (that we know about), of those initials in the Library and Information Science (CIG) sector. We may have been a bit quiet, or at least not on the radar of many readers of this journal, so I welcome the invitation to explain what we are doing these days.

Actually we are still doing what we always did in the field of knowledge management, covering everything from indexing, to metadata (and the difference is …?), to classification, to the semantic web, and probably stuff I have forgotten at the time of writing.

As a Special Interest Group (SIG) of CILIP we have to produce an annual activity plan and the areas I shall describe are largely in line with the latest version of that document: continuing professional development, standards, good practice and networking/ communities of practice.

Continuing Professional Development

Opportunities to attend events to enhance one's professional development might be seen as one of the foremost purposes for belonging to any professional group.

https://doi.org/10.1017/S1472669610000265 Published online by Cambridge University Press