Knowl. Org. 32(2005)No.2
Book Reviews

95

References

Bowker, G.C. and S.L. Star. 1999. *Sorting things out: classification and its consequences.* Cambridge: MIT Press.

Foucault, M. 1979. *Discipline and punish: the birth of the prison*, tr. Alan Sheridan. Vintage Books, pp. 195–230.

Rasmussen, J., A.M. Pejtersen and L.P. Goodstein. 1994. *Cognitive systems engineering.* New York: Whiley.

Szostak, R. 2003. Classifying scholarly theories and methods. *Knowledge organization* 30:20–35.

ROE, Sandra K. and Alan R. THOMAS (Editors). *The Thesaurus: Review, Renaissance and Revision.* Binghamton: The Haworth Information Press, 2004. 209 pp. ISBN 0-7890-1978-7 (hardcover) 0-7890-1979-5 (paperback).

It might be thought unfortunate that the word thesaurus is assonant with prehistoric beasts but as this book clearly demonstrates, the thesaurus is undergoing a notable revival, and we can remind ourselves that the word comes from the Greek *thēsauros,* meaning a treasury. This is a useful and timely source book, bringing together ten chapters, following an Editorial introduction and culminating in an interview with a member of the team responsible for revising the NISO Standard *Guidelines for the construction, format and management of monolingual thesauri*; formal proof of the thesaural renaissance.

Though predominantly an American publication, it is good to see four English authors as well as one from Canada and one from Denmark; and with a good balance of academics and practitioners. This has helped to widen the net in the citing of useful references. While the techniques of thesaurus construction are still basically sound, the Editors, in their introduction, point out that the thesaurus, in its sense of an information retrieval tool is almost exactly 50 years old, and that the information environment of today is radically different. They claim three purposes for the compilation:

> "to acquaint or remind the Library and Information Science community of the history of the development of the thesaurus and standards for thesaurus construction.

> to provide bibliographies and tutorials from which any reader can become more grounded in her or his understanding of thesaurus construction, use and evaluation.

> to address topics related to thesauri but that are unique to the current digital environment, or network of networks."

This last purpose, understandably, tends to be the slightly more tentative part of the book, but as Rosenfeld and Morville said in their book *Information architecture for the World Wide Web* "thesauri [will] become a key tool for dealing with the growing size and importance of web sites and intranets". The evidence supporting their belief has been growing steadily in the seven years since the first edition was published.

The didactic parts of the book are a collection of exercises, readings and resources constituting a "Teach yourself" chapter written by Alan Thomas, ending with the warning that "New challenges include how to devise multi-functional and user-sensitive vocabularies, corporate taxonomies and ontologies, and how to apply the transformative technology to them." This is absolutely right, and there is a need for some good writing that would tackle these issues. Another chapter, by James Shearer, skilfully manages to compress a practical exercise in building a thesaurus into some twenty A5 size pages. The third chapter in this set, by Marianne Lykke Nielsen, contains extensive reviews of key issues and selected readings under eight headings from the concept of the thesaurus, through the various construction stages and ending with automatic construction techniques.

Leonard Will addresses the topic of thesaurus consultancy with sensible remarks that relate to any sort of consultancy. He points out that the phrase 'thesaurus consultancy' is probably too narrow, as there is much deep investigation to be conducted, and decisions to be made in definition and planning before the actual building and implementation of any structured vocabulary should begin. Too many clients seem to believe that a thesaurus is some back-room tool to be built and plugged in when ready. In fact, the thesaurus (or any other authority file providing the values for metadata) should be an integral part of the overall information architecture, and its compilation should not be regarded as a trivial task. Will quotes some cost figures to support this view. In the following chapter, Leslie Ann Owens and Pauline Atherton Cochrane suggest that the seminal *Thesau-*

96

Knowl. Org. 32(2005)No.2
Book Reviews

rus of Engineering and Scientific Terms probably cost almost $1 million dollars in 1967 dollars…equivalent to about $40 dollars per term. However, they conclude that today, a realistic figure is more likely to be "$10 a term if all computer programs and input procedures are optimal". This is a very useful and important chapter, which reviews thesaurus evaluation throughout its life cycle, a vital exercise if the thesaurus is to be properly maintained and updated. Many thesauri have been allowed to fall into such disrepair that they cease to be effective.

Melissa Riesland provides another useful and succinct chapter, this time on vocabulary management software, mentioning taxonomies and topic maps *en passant*, and citing Leonard Will's useful website where a wide range of these tools are analysed and details of their characteristics and functionalities are presented. Jane Greenberg gives a rather discouraging but accurate account of user comprehension, quoting figures from a study which found that participants' thesaurus knowledge was extremely low: only 14.3% saying that they had used a thesaurus for searching before participating in the study. This underlines the need for user training, which should form part of a wider approach to the whole aspect of information literacy. Greenberg quotes work by Bates on the 'end-user thesaurus,' which attempts to capture and structure the uncontrolled terms used by searchers. Since this book was written there has been much debate concerning the efficacy of "folksonomies," collaborative vocabularies compiled by individuals who separately tag material of interest, thereby building a sort of consensus over time. There are several instances of such web-based social networks, which are likely to become more popular, and thesaurus compilers should take account. In line with the international flavour of the book, Patrice Landry looks at the multilingual problem. This chapter is mainly concerned with a review of MACS (Multilingual Access to Subjects); a project with the strategy of developing a Web-based link and search interface through which equivalents between three Subject Heading Languages can be created and maintained, and by which users can access online databases in the language of their choice. The three systems in the project are German, French and English language. With the dramatic spread of use of the Web, particularly in the Far East, such projects are going to be increasingly valuable and important. Continuing the theme of networking, Eric Johnson discusses distributed thesaurus web services within the near-anarchy of the World Wide Web. Johnson points out,

rightly, that "HTML-based information services provide access to online information sources but do not make them useful for much more than viewing in a web browser." There is also no cohesive cataloguing or subject access scheme for the Internet. Addressing this problem, Johnson proposes the Thesauro-Web, a network of thesaurus access and navigation services using XML, and this idea would seem to be a necessary adjunct to the Semantic Web.

The interview with Amy Warner is, inevitably, out of date. The document *Z39.19,* now retitled *Guidelines for the construction, format and management of monolingual structured vocabularies* has now been published in its consultation edition. This retitling may be to accord with a recommendation of the 1999 *Workshop on Electronic Thesauri,* quoted by Jean Aitchison and Stella Dextre Clarke in their chapter: "The standard should provide for a broader group of controlled vocabularies than those that fit the standard definition of the 'thesaurus'. This includes, for example, ontologies, classifications, taxonomies and subject headings, in addition to standard thesauri. The primary concern is with shareability (interoperability) rather than with construction or display. Therefore this new standard will probably not supersede Z39.19, but supplement it". Meanwhile, in the U.K. a small Working Group, which has exchanged ideas with its American colleagues, has been revising the old *BS 5732; Guidelines for the establishment and development of monolingual thesauri* (which was adopted as its ISO equivalent ISO 2788), and *BS 6723 Guidelines for the establishment and development of multilingual thesauri.* This is a more ambitious undertaking, which not only does address the various manifestations of structured vocabularies (thesauri, taxonomies, subject headings, classification – including business classification for file plans, and ontologies), but will be presented in five parts under the new title *BS 8723 Structured vocabularies for information retrieval.* The parts, the first two of which have been published for consultation, are:

Part 1: Definitions, symbols and abbreviations
Part 2: Thesauri
Part 3: Vocabularies other than thesauri (provisional title)
Part 4: Interoperability between vocabularies (provisional title)
Part 5: Protocols and formats needed for exchange of vocabulary data (provisional title).

Knowl. Org. 32(2005)No.2
Book Reviews

97

Finally, though it is the first chapter in the book, Jean Aitchison and Stella Dextre Clarke make three extremely pertinent points in their "look to the future":

> We face a paradox. Ostensibly, the need and the opportunity to apply thesauri to information retrieval are greater than ever before. On the other hand, users resist most efforts to persuade them to apply one.
> The drive for interoperability of systems means we must design our vocabularies for easy integration into downstream applications such as content management systems, indexing/meta-tagging interfaces, search engines, and portals.
> Summarizing the search for vocabularies that work more intuitively, we see that there are trends working in opposite directions. In the hugely popular taxonomies on the one hand, relationships between terms are more loosely defined than in thesauri. In the ontologies that will support computer-to-computer communications in AI applications such as the Semantic Web, we see the need for much more precisely defined term relationships.

This is a useful and approachable book. It is a pity that the index is such a poor advertisement for vocabulary control and usefulness.

Alan Gilchrist

Alan Gilchrist, Cura Consortium and TFPL Ltd.