116

Knowl. Org. 31(2004)No.2
Book Reviews

Knowledge Organization and International Information Retrieval. Edited by Nancy J. Williamson and Clare Beghtol. *Cataloging & Classification Quartely* 37, 1/2 2003. ISSN 0163-9374.

This is a collection of 14 papers covering a broad range of issues relating to knowledge organization and classification and their current status in new information environments, with a particular slant towards multilingual and multicultural contexts. As mentioned in the introduction, authors were given the mandate to discuss knowledge organization issues in the context of international, multilingual, multicultural and multi-scheme information environment. This reviewer believes this aim has been achieved considering the breath of topics and issues covered in the papers, ranging from classification schemes and thesauri to text mining, natural language processing and metadata mapping. Putting knowledge organization in the context of the web, linguistic research, information retrieval and knowledge discovery has provided a ground for researchers to gain an insight into the scale and scope of research and development in the area.

The articles have been categorised under four main themes. The first section addresses *General bibliographic systems*, and features the current and future applications in electronic environments of general classification schemes such as the Dewey Decimal Classification (DDC), the Universal Decimal Classification (UDC) and the Library of Congress Classification (LCC). Some of the main features of this section are: a comparison of the use of general and special classification systems, a demonstration of the unsuitability of general classification systems to represent documents in a special domain, discussions of how to adapt and map general classification schemes such as DDC to specialist thesauri, discussions of how to make translated classification schemes more hospitable, a description of problems associated with translation from one language into another, and a discussion of the extent to which a UDC-based multilingual thesaurus can affect information retrieval performance.

The theme of the second part of this collection is *Information organization in knowledge resources*. Two papers address the issues of subject approach, subject-based retrieval and the internationalization of knowledge organization in the context of the web. In particular, the complexity of subject access in a wider context with a variety of subject access systems is highlighted. There is also a discussion of the ways in which education resources on the web can be sought and retrieved through identifying and comparing subjects in the area of education in classification schemes such as DDC, UDC and LCC. Two papers focus on knowledge discovery techniques such as text mining, data mining, and text analysis to improve retrieval performance. There is a useful discussion of text mining and its importance in the web environment. Four layers of text mining have been outlined, namely linguistic analysis, co-occurrence analysis, clustering and categorizations and visualization. In addition, discussions of issues such as the use of knowledge discovery techniques in bibliographic databases and cross-links between and among classification schemes and thesauri suggest ways in which retrieval can be improved.

Linguistics, terminology and natural language processing are at the centre of the third section. Human language technologies in information society, the issue of information access across language boundaries, the role of human language technologies in information access, natural language processing, ontologies, text summarization, machine translation, cross-language information retrieval, linguistic approaches to knowledge organization such as semantic and lexical analysis are among the topics discussed. A most interesting contribution reports on how end-users view and relate metadata elements to their definitions in terms of clarity and understandability. The paper stresses the importance of reviewing and refining terminology and definition in the light of optimising users' access to knowledge repositories without users having to know much about individual metadata schemes.

The final section, titled *Knowledge in the world and the world of knowledge* consists of three papers. Readers of this section are introduced to the theoretical and philosophical bases of knowledge organization and management through a theoretical comparison of various classification schemes. There is a discussion of ontologies, of ontological bases for knowledge organization, and of semantic integration and sharing of ontologies. Also discussed are the value and importance of tacit and explicit organizational knowledge and knowledge generation and transaction in organizations. The final paper compares some theoretical features of two classification schemes, namely the North American Industry Classification System and LCC, with respect to their coverage of geographic and international economic subject categories.

The articles in this collection provide a well-balanced account of recent developments and chal-

Knowl. Org. 31(2004)No.2
Book Reviews

117

lenges of knowledge organization with a view to globalization and internationalization issues. This book will be a very useful source for those researchers who are interested in the ways in which knowledge organization systems have been used and are being used in different information contexts. It is recommended reading for researchers in the areas of knowledge organization, information retrieval, and language processing, as well as for students taking courses on advanced topics in information and knowledge organization.

A. Shiri

Dr. Ali Shiri, School of Library and Information Studies, University of Alberta, Edmonton, Alberta T6G 2J4 Canada. Email: ashiri@ualberta.ca

Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001 and sponsored by the IFLA Classification and Indexing Section, the IFLA Information Technology Section and OCLC. Ed. I.C. McIlwaine. München: K.G. Saur, 2003. ix, 193 p. ISBN 3-598-11634-9.

This excellent volume offers 22 papers delivered at an IFLA Satellite meeting in Dublin Ohio in 2001. The conference gathered together information and computer scientists to discuss an important and difficult question: in what specific ways can the accumulated skills, theories and traditions of librarianship be mobilized to face the challenges of providing subject access to information in present and future networked information environments?

The papers which grapple with this question are organized in a surprisingly deft and coherent way. Many conferences and proceedings have unhappy sessions that contain a hodge-podge of papers that didn't quite fit any other categories. As befits a good classificationist, editor I.C. McIlwaine has kept this problem to a minimum. The papers are organized into eight sessions, which split into two broad categories. The first five sessions deal with subject domains, and the last three deal with subject access tools.

The five sessions and thirteen papers that discuss access in different domains appear in order of in-

creasing intension. The first papers deal with access in multilingual environments, followed by papers on access across multiple vocabularies and across sectors, ending up with studies of domain-specific retrieval (primarily education). Some of the papers offer predictably strong work by scholars engaged in ongoing, long-term research. Gerard Riesthuis offers a clear analysis of the complexities of negotiating non-identical thesauri, particularly in cases where hierarchical structure varies across different languages. Hope Olson and Dennis Ward use Olson's familiar and welcome method of using provocative and unconventional theory to generate meliorative approaches to bias in general subject access schemes. Many papers, on the other hand, deal with specific ongoing projects: Renardus, The High Level Thesaurus Project, The Colorado Digitization Project and The Iter Bibliography for medieval and Renaissance material. Most of these papers display a similar structure: an explanation of the theory and purpose of the project, an account of problems encountered in the implementation, and a discussion of the results, both promising and disappointing, thus far. Of these papers, the account of the Multilanguage Access to Subjects Project in Europe (MACS) deserves special mention. In describing how the project is founded on the principle of the equality of languages, with each subject heading language maintained in its own database, and with no single language used as a pivot for the others, Elisabeth Freyre and Max Naudi offer a particularly vivid example of the way the ethics of librarianship translate into pragmatic contexts and concrete procedures.

The three sessions and nine papers devoted to subject access tools split into two kinds: papers that discuss the use of theory and research to generate new tools for a networked environment, and those that discuss the transformation of traditional subject access tools in this environment. In the new tool development area, Mary Burke provides a promising example of the bidirectional approach that is so often necessary: in her case study of user-driven classification of photographs, she uses personal construct theory to clarify the practice of classification, while at the same time using practice to test the theory. Carol Bean and Rebecca Green offer an intriguing combination of librarianship and computer science, importing frame representation technique from artificial intelligence to standardize syntagmatic relationships to enhance recall and precision.

The papers discussing the transformation of traditional tools locate the point of transformation in dif-