

Relationships in the Organization of Knowledge

Carol A. Bean and Rebecca Green (editors)
(University of Tennessee and University of Maryland)

Dordrecht: Kluwer Academic
Publishers (Information science and
knowledge management series, edited
by J. Mackenzie Owen, volume 2), 2001,
ix+232 pp; hardbound, ISBN
0-7923-6813-4, \$97.00, £61.00, €90.00

Reviewed by
Gregory Grefenstette
Clairvoyance Corporation

This book is not for linguists. This book is not for computer scientists. This book is for undergraduate librarians.

Given the promising title, from which one might expect a comprehensive look at imposing order on knowledge, this is a little disappointing. The back-cover blurb insists that the book can be used to provide “guidance for relational tasks [that are] now taking on greater significance, as retrieval systems increasingly operate in automated modes and as retrieval systems cross linguistic, cultural, and disciplinary boundaries.” But going inside, one enters the closed-world community of library science for which the automation of the card catalogue is still big news.

The book is disappointing because so many strands of modern computational and linguistic work attacking the problem of organizing knowledge are missing. For example, from computer science, the fact that relationships have formed the core of database modeling ever since relational databases (Codd 1970) and the entity-relationship model (Chen 1976) goes unacknowledged. From artificial intelligence, missing is any of the work on ontologies for organizing knowledge, such as Penman’s upper model (Bateman et al. 1990) and CYC (Lenat et al. 1990), and work on organizing knowledge in expert systems. As for linguistics, there is no mention of the strain of research stemming from the idea of semantic markers (Katz and Fodor 1963) and the myriad subsequent efforts to exploit them to represent meaning despite their inherent limitations (Bolinger 1965; Eco 1976). All this practical and intellectual effort is ignored here. The content of this book is light years away from current practical concerns of knowledge representation (www.kr.org), knowledge management (www.cikm.org), or the relationships within domains of knowledge being explored with UML (www.uml.org) and XML (www.xml.org). Maybe these criticisms all derive from false hopes raised by the book’s title. A better title for this book would be *Relationships in the Organization of Library Books*.

The first part of the book is “Theoretical Background.” Rebecca Green opens the book by describing the wide variety of efforts to characterize relations from the bibliographic and documentary points of view. Barbara Tillet follows with a brief, low-level tutorial on how bibliographic items can be related to each other—for example, how to describe “works within works,” in which one book contains other books. Next, Stella Dextre Clarke describes the International Standards Organization (ISO) relations used in manually built thesauri (e.g., BT for broader term, NT for narrower term, and their multiple variants). Here, Dextre Clarke shows how the same word is described incompatibly in different thesauri; she also shows in a table a number of ways in which the

RT (related term) relationship actually covers a wide variety of relations (e.g., “whole-part,” “action and its patient,” “causal dependence,” “concept and its opposite,” and 11 other relations with examples). Jessica Milstead, in her chapter about standards for relations between subject-indexing terms, laments how much latitude is possible in interpreting relations mentioned in thesaural standards. Their interpretations, she says, may be reasonably evident to a human manipulating the thesaurus, but she warns that “[i]t is when the thesaurus sits behind the interface as a tool to be applied [automatically] by the system that more precise specification of equivalence relations is required”.

Michele Hudon describes the drawbacks of multilingual thesauri, so that, for example, even though words may pass as translations, restriction of meanings makes thesaural relations between near-translations quickly diverge. For example, translations of the hyponyms of the English word *education* are not hyponyms under the French word *éducation*, which, though a near-translation, has a more restricted meaning than the English word. Some of the translations would have to be hyponyms of the French word *enseignement*. At the end of her contribution, Hudon refers to work on EuroWordNet (Vossen 1998) that has been addressing this problem.

Olivier Bodenreider and Carol A. Bean, one of the editors, talk about some of the problems involved in integrating different thesauri into one coherent organization, giving details about the structure of shared medical vocabularies of the Unified Medical Language System (UMLS). This chapter contains many pointers to medical terminological resources on the Web, to which one could add the International Medical Informatics Association Working Group 6 on Medical Concept Representation (www.mayo.edu/imia-wg6/).

Clare Beghtol offers a philosophical musing on structure and meaning, following a historical trajectory from Aristotle through to the Dewey decimal system and social psychology. The editors then come back with a chapter on relevance relations. This chapter might be at home in a book about information retrieval (which is adequately cited here), but seems out of sync with the rest of the book, which hovers around classic cataloguing and thesaurus use. In this chapter, Bean and Green present some of their previous research that attempted to typify the relations between a topical guide to scriptures and the passages cited as relevant to the topics in the guide. The research found that “topical relevance relations include a wide variety of relationships, only some of which, perhaps only a relatively small proportion, are matching relations” (page 127). Considering that matching relations are what people use now to browse the Web, this conclusion is interesting, as is the rest of this chapter’s discussion about what relevance means.

The second part of the book, titled “Systems,” contains chapters reviewing Library of Congress subject headings, the structure of the Art and Architecture Thesaurus (www.getty.edu/research/tools/vocabulary/aat/), Medical Subject Headings (MeSH, www.nlm.nih.gov/mesh/), a multicultural and multilingual thesaurus from India, Colon Classification (an attribute-value-type library classification system developed by S. R. Ranganathan in the early 1950s), and the Dewey decimal system. This part of the book might be useful to a computational linguist who needs a quick reference to one of the classic classification systems devised for organizing library stocks.

References

Bateman, John, Robert Kasper, Johanna Moore, and Richard Whitney. 1990. A general organization of knowledge for

natural language processing: The Penman upper model. Technical Report, Information Sciences Institute, Marina del Rey, CA.

- Bolinger, Dwight. 1965. The atomization of meaning. *Language*, 41(4):555–573.
- Chen, Peter Pin-Shan. 1976. The entity-relationship model: Towards a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36.
- Codd, E. F. 1970. A relational model for large shared data banks. *Communications of the ACM*, 13(6):377–387.
- Eco, Umberto. 1976. *A Theory of Semiotics*. Indiana University Press, Bloomington.
- Katz, Jerrold J. and Jerry A. Fodor. 1963. The structure of a semantic theory. *Language*, 39(2):170–210.
- Lenat, Douglas B., Ramanathan V. Guha, Karen Pittman, Dexter Pratt, and Mary Shepherd. 1990. CYC: Toward programs with common sense. *Communications of the ACM*, 33(8):31–49.
- Vossen, Piek. 1998. *EuroWordNet*. Kluwer, Dordrecht.

Gregory Grefenstette, Principal Research Scientist at Clairvoyance Corporation, is the author of *Explorations in Automatic Thesaurus Discovery* (Kluwer Academic Publishers, 1994) and editor of *Cross-Language Information Retrieval* (Kluwer Academic Publishers, 1998). He has reviewed manuscripts for many natural language processing, artificial intelligence, and computational linguistics journals and conferences. He is currently working on very large lexicons. Grefenstette's address is Clairvoyance Corporation, 5301 Fifth Avenue, Pittsburgh, PA 15232; e-mail: g.grefenstette@Clairvoyancecorp.com.