180

Knowl. Org. 27(2000)No.3
Book Reviews

on the generation of summaries from event data. Maybury presents SumGen, a system that selects key information from an event database based on frequencies about events and relations, and domain importance measures.

The section on "Evaluation Methods," is concerned with the fundamental problem of assessing the quality and success of automatic abstracts. The first paper, by Rath, Resnick, and Savage, presents two experimental studies of abstracting procedures. The first study measures the agreement between humans and algorithms in the task of selecting representative sentences whilst the second tests the reliability of humans in preparing abstracts by selecting representative sentences. The following paper, by Brandow, Mitze and Rau, presents an evaluation of ANES (Automatic News Extraction System), a domain-independent summarization system of news documents that uses a combination of statistic and heuristic methods. The authors show that a simple system that selects the leading part of texts outperforms ANES in an acceptability evaluation. The section continues with a paper by Morris, Kasper, and Adams that investigates the effects and theoretical limitations of automatic extracts in effective message comprehension. Firmin and Chrzanowski conclude the section with their description of a dry run of an extrinsic, task-based evaluation of automatic summarization systems as part of the TIPSTER program of the Defense Advanced Research Projects Agency (DARPA).

The last section, "New Summarization Problem Areas," contains five papers dealing with multi-document and multimedia summarization. Salton, Singhal, Mitra, and Buckley use ideas from inter-document link generation to produce intra-document links between various paragraphs and sentences of single articles on the basis of vocabulary overlap. The text structure so produced is used to select passages that constitute an extract of the source document. Various orders of the text structure are explored in order to enforce coverage and coherence of the extract. In the second paper of this section, Mani and Bloedorn explore the use of cohesion relations between proper names to construct user-focused summaries of multiple articles. In their approach, cohesion is first used to construct a text graph representation of each source document. This graph contains nodes representing terms in the source document and links representing cohesion relations. A spreading activation algorithm is then used to find nodes related to terms in a user query which is the input to the sys-

tem. The authors describe their algorithm for finding similarities and differences among text segments in different sources and illustrate a variety of presentation strategies. In the next paper, McKeown and Radev present SUMMONS, a system that summarizes a series of news articles on the same event. The input to the system is a set of instantiated templates obtained from an Information Extraction system. These templates contain the salient facts of texts on terrorism. In their approach, domain dependent summary operators for linking information from different templates are used for content planning.

The final two papers are less related to the theme of text summarization. Merlino and Maybury report on an empirical evaluation of the Broadcast News Navigator System that uses linguistic and graphical elements to support detection and extraction of information from broadcast news. Their experiment measures how mixed representation methods influence the identification and comprehension tasks. The last paper, by Futrelle, is an initial exploration of issues involved in diagram summarization. The author identifies issues related with selection, simplification, merge and generation of diagrams.

In summary, this is a well organized volume, with a two column index providing access to specific topics and terminology. Most papers include figures, tables and good bibliographic references. While the book does not cover all the relevant research in the field and includes some articles on the edge of text summarization, it is certainly a timely document recommended for all interested in text summarization, information retrieval, computational linguistics, and related topics. It is also a good source of documentation for the preparation of seminars and tutorials.

Horacio Saggion

Dr. H. Saggion, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, S1 4DP, Sheffield, UK. E-mail: h.saggion@dcs.shef.ac.uk

BRENNER, Diane, and ROWLAND, Marilyn (eds.). **Beyond Book Indexing : How to Get Started in Web Indexing, and Other Computer-Based Media**. Phoenix, AZ : American Society of Indexers / Information Today, 2000. 149 p. ISBN 1-57387-081-1.

This book is divided in four sections, each presenting and analyzing some basic aspects of indexing: em-

Knowl. Org. 27(2000)No.3
Book Reviews

181

bedded indexing, Web indexing, computer-based indexing, and marketing strategies on the Web for indexers. Each section includes two to three articles written by professional indexers who present their point of view or personal experience on the subject discussed and, if appropriate, guide readers with a step-by-step approach to a particular indexing process. Each article is accompanied by a list of resources (articles, books, and online references) directing readers to more information on the subject. A glossary listing terms which may be unfamiliar to readers and a complete index are also included at the end of the book.

The first section of the book, entitled "Beyond Stand-Alone Indexes: Embedded Indexing," gathers two articles on this subject; it is more likely to be of interest to beginners than to experienced indexers, since the latter are often already producing embedded indexes. The first article provides an interesting overview of embedded indexing (definition, process, hardware and software, pricing, etc.) but suffers from a lack of practical examples which would help readers to visualize the work involved in this type of indexing. The second article discusses the specific aspects of indexing computer-related documents, and draws a comparison with more traditional subject fields.

Indexers interested in Web indexing will appreciate the second section of the book, "Beyond the Book: Web Indexing," and more specifically the first two articles written respectively by Dwight Walker, and Kevin Broccoli and Gerry Van Ravenswaay. These essays summarize the origins, evolution, types and numerous challenges of Web indexing. They also explain the difficulties and limitations of the indexing process with HTML, compare Web indexes to search engines, briefly describe HTML codes and give readers tips on graphic design and layout of online indexes. A well-developed and very complete article by Seth Maislin concludes this section, with a description of a step-by-step approach to the process of creating a Web site index for an online newsletter.

The third part of the book, "Special Topics in Computer-Based Indexing," discusses the usefulness and importance of "description" and "keywords" meta tags in Web sites, as well as the process of indexing multimedia CD-ROM and Windows-based online help. Texts are completed by several examples illustrating the authors' comments.

The first and second articles of the last section of the book, "Beyond Traditional Marketing : Selling Yourself in Hyperspace," demonstrate the usefulness of a website as a marketing tool for indexers and the information contained in such a site. They give indexers tips on the "do's and don'ts" of contents, graphics and layout for their own website, and present an up-to-date information on tools available to set up or promote a website. The third article, by Dan Connolly, defines the concept of list servers and explains the specificities of announcement lists and discussion lists. Connolly also provides information on the creation of a personal email discussion list, covering subjects such as servers, software, management, policies and procedures, publicity, fostering discussion, mediation and privacy.

This book will primarily interest beginners in indexing or experienced indexers not quite familiar or up-to-date with new information technologies and Web indexing. It is fair to say that the main purpose of this book seems to give readers a broad overview or description of the subject dealt with by the authors. Articles rarely present an in-depth analysis of the subject discussed but do direct readers to extensive references for further information.

Ginette Chandonnet

G. Chandonnet, Librarian and freelance indexer, Montréal, Canada. E-mail: soumah@istar.ca.