

Briefly Noted

Automatic Indexing and Abstracting of Document Texts

Marie-Francine Moens

(Katholieke Universiteit Leuven)

Dordrecht: Kluwer Academic Publishers (The Kluwer international series on information retrieval, edited by W. Bruce Croft), 2000, xvi+265 pp; hardbound ISBN 0-7923-7793-1, \$110.00, £76.00, Dfl 260.00

With the explosion in the quantity of on-line information in recent years, automatic abstracting and indexing has received renewed interest and a number of promising approaches have emerged. The goal of this book is to present a complete description of current indexing and abstracting techniques in the context of the underlying linguistic and statistical knowledge.

The book has three parts: the indexing and abstracting environment, methods of automatic indexing and abstracting, and applications. The first part covers theories about text, such as rhetorical and thematic structure, and discusses the use of different text representations for information retrieval and abstraction. The second part of the book deals with lexical analysis and weighting schemes, abstraction, and evaluation techniques. The third part describes two summarization systems developed by the author: a summarizer for court cases and a generic magazine article summarizer.

I found the application part particularly interesting because it describes all the stages of the development of applied summarization systems, from the initial corpus analysis to the final evaluation of the method. The system developed by the author combines symbolic techniques, such as a text grammar, with statistical methods, such as classification and clustering. The system demonstrates tradeoffs between symbolic techniques based on manual knowledge encoding and corpus-based shallow methods.

However, the book's breadth comes at the expense of its depth. This problem is especially acute in the first two parts of the book, which are supposed to give an overview of the subject. The presentation of text theories and abstraction techniques is somewhat superficial and incomplete; the book is not up to date and does not contain the latest developments in the area. For exam-

ple, it omits computational methods for discourse; cohesion-based approaches; segmentation algorithms; and recent work in evaluation, corpus-based summarization, and regeneration (e.g. Marcu 1997; Mani et al. 1998; Jing and McKeown 1999). These limitations can be a major hurdle for the reader who wants to learn about state-of-the-art developments.—*Regina Barzilay, Columbia University*

References

- Mani, Inderjeet, David House, Gary Klein, Lynette Hirschman, Leo Obrst, Therese Firmin, Michael Chrzanowski, and Beth Sundheim. 1998. The TIPSTER SUMMAC Text Summarization Evaluation, Final Report. The MITRE Corporation, report number MTR 98W0000138.
- Marcu, Daniel. 1997. From discourse structures to text summaries. *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization*, pages 82–88, Madrid, Spain.
- Jing, Hongyan and Kathleen McKeown. 1999. The Decomposition of Human-Written Summary Sentences. *Proceedings of the 22nd International Conference on Research and Development in Information Retrieval (SIGIR'99)*. Berkeley.

Handbook of Multimodal and Spoken Dialogue Systems: Resources, Terminology and Product Evaluation

Dafydd Gibbon, Inge Mertins, and Roger K. Moore (editors)

(University of Bielefeld, DERA, and 20/20 Speech Ltd.)

Boston: Kluwer Academic Publishers (The Kluwer international series in engineering and computer science, volume 565), 2000, xix+519 pp and CD-ROM; hardbound, ISBN 0-7923-7904-7, \$175.00, £123.00, Dfl 440.00

"Dictation systems, read-aloud software for the blind, speech control of machinery, geographical information systems with speech input and output, and educational software with 'talking head' artificial tutorial agents are already on the market. The field is expanding rapidly, and new methods and applications emerge almost daily. But good sources of systematic information have not kept pace with the body of information