quite familiar to the community of researchers, will continue to be described as non-conventional! Section H (chapter 14) discusses the evaluation of indexing systems in traditional terms of measuring relevance and recall (a printing error describes both of these ratios by the same equation). Section I (the last chapter) presents a few of the major international indexing and abstracting services.

The inherent objective of the author is to provide Indian students with a well illustrated textbook on subject indexing. But unfortunately, the content is dated and it hardly addresses indexing needs and techniques in a networked information environment. The book emphasizes historical developments more than current issues. References appended to most of the chapters are old. In the selective list of thesauri (p. 61) the most recent entry dates back to 1981. Nevertheless, some of the systems described are treatises in terms of details.

At best, this book can be viewed as a record of India's contribution to the art and science of indexing. Indexing is viewed here in the narrow sense of concept formation and their arrangement rather than in the wider sense of it being instrumental to navigating a text or a collection of documents, or as a tool for organisation. There is no discussion of classification as indexing tool. Categorizing classification as "grouping" and associative is shallow and misleading. Can we group objects without identifying their inter and intra group associations? All classifications in any form are associative and correlative. There is no chapter on subject headings lists even though they are the most widely used tools for vocabulary control and subject indexing in library catalogues.

The book has been meticulously edited, notwithstanding some typos here and there (Mortimer Taube always appear as Martimer Taube for example). The text of each chapter has been divided into sections with number and feature headings, and ends with a summary or conclusion. The style of references is uniform, as per Indian standards.

For Indian students, this is a handy one-stop shop for the major traditional subject indexing systems of the world. For the readers abroad, it is handy to understand Indian subject indexing methods *sans* facet analysis and the CC.

M.P. Satija

Dr. M.P. Satija, GND University, Amritsar, 143 005 India. E-mail: dsce_gndu@yahoo.com or manmohan@myhost.gndu.ernet.in.

MANI, Inderjeet, and MAYBURY, Mark T., eds. **Advances in Automatic Text Summarization**. Cambridge, MA : MIT Press, 1999. 434p. ISBN 0-262-13359-8.

The 26 papers in this book capture some of the most relevant research and development in automatic text summarization, which is the process of distilling the important information from a source document to produce an abridged version. Thirteen of these articles are new while the remaining essays are reprints from books, journals, or conference proceedings. The volume is organized into an introduction providing terminological details, an overview of the content, and pointers to useful resources with related information. The introduction is followed by a position paper by Karen Spärk Jones. The rest of the volume is developed into six sections representing key areas of research and development in text summarization: "Classical Approaches," "Corpus Based Approaches," "Exploiting Discourse Structure," "Knowledge-rich Approaches," "Evaluation Methods," and "New Summarization Problem Areas." Each section starts with a rich and well organized introduction which includes a summary of each paper and references to additional bibliographic material.

In her paper, Karen Spärk Jones suggests that progress in automatic summarization demands a better, more focused research methodology. Having introduced a basic three-stage process model of text summarization, she describes the current state of affairs in the field, and concentrates on context factors that affect summarization. Spärk Jones suggests that, as we cannot expect computers to emulate humans in the production of summaries, the nearer-term strategy should be on shallow processing.

The first section, "Classical Approaches," contains three reprints of journal articles going back to the very foundations of automatic text summarization. Luhn's paper describes the first implemented sentence extraction algorithm which uses term frequencies to measure sentence relevance. Luhn's algorithm filters terms using a stop-list, and computes term frequencies by aggregating terms on the basis of orthographic similarity. These term frequencies are later used to score and select sentences for the abstract. The next paper, by Edmunson, studies how combinations of different linguistic and structural features affect the co-selection ratios between automatic abstracts and ideal abstracts. Edmunson investigates the presence of pragmatic words (cue method), title and heading words (title method), and structural indicators (loca-

tion method), as additional criteria for sentence worthiness. The author demonstrates that a combination of cue, title and location methods produces the highest mean co-selection score. In the last paper of the section, Pollock and Zamora describe the Automatic Document Abstracting Method (ADAM) used at Chemical Abstracts Service (CAS) to produce indicative abstracts which conform to CAS standards. ADAM is based on the idea of sentence rejection using cue words. ADAM automatically edits sentences for abbreviation and compaction.

"Corpus-based Approaches," are covered in four papers concerned with the issue of how different textual features can be extracted from text corpora and manually or automatically combined to produce better abstracts. The first paper, by Kupiec, Pedersen and Chen, considers text summarization as a statistical classification problem. The authors propose, for the first time, a Bayesian classifier, trained on a set of source documents/extracts pairs, that estimates sentence worthiness using features such as sentence length and sentence location. The next two papers describe applications of Kupiec, Pedersen and Chen's work. Myaeng and Jang present a sentence extraction system for technical texts in Korean; they consider other features besides those presented by Kupiec et al., and combine them using Dempster-Shafer Theory. Aone, Okurowski, Gorlinsky, and Larsen, describe DimSum, a sentence extraction system which uses text statistics and corpus statistics to derive signature words as one feature for text summarization. In the last paper of the section, Hovy and Lin focus on topic identification and fusion for text summarization. With regards to topic identification, they use a new algorithm for the automatic identification of sentence positions carrying important topics. Further, they manually combine sentence position with other relevant features to score sentences. On topic fusion, Hovy and Lin explore concept counting using the WordNet lexical database, text categorization, and text clustering.

The third section, "Exploiting Discourse Structure," explores, in five uneven papers, the global properties of the text such as cohesion, coherence and rhetorical relations. The first paper, by Boguraev and Kennedy, focuses on the linguistic processes underlying the automatic identification of topic stamps, i.e., phrasal units that represent the document's content using an anaphora resolution algorithm. Text summaries are produced by presenting topic stamps in appropriate sentence contexts. In the second paper, Barzilay and Elhadad propose lexical chains for text interpretation, i.e., sequences of sentences grouped together by cohesion relations found in WordNet. The authors explore different heuristics for selecting sentences from lexical chains in order to produce text summaries. Daniel Marcu, in his paper, describes an interesting psycholinguistic experiment that shows that the concept of discourse structure and nuclearity can effectively be used in text summarization. He proposes and evaluates a novel automatic discourse-based summarization algorithm that uses the output of a rhetorical parser to construct text summaries. The fourth paper, by Strzalkowski, Stein, Wang, and Wise, presents a method for summarization of news articles based on the selection of paragraphs referring to 'what is new' in the text. Paragraphs dealing with background information are selected as well in order to improve coherence. The section closes with an excellent work by Teufel and Moens in which they exploit the argumentative structure of scientific texts as a means to construct flexible abstracts. Their sentence extraction system is based on Bayesian classifiers: it first extracts abstract-worthy sentences and then classifies them in rhetorical roles. The authors argue that superficial features of the text can be effectively used for this task.

The fourth section, "Knowledge-rich Approaches," contains four papers focusing on summarization of genre and domain specific texts. Lehnert's article provides the foundation for narrative summarization through the development of the theory of plot units, i.e., configurations of affect states found in narratives. The author argues that narrative texts can be interpreted as configurations of plot units which can be effectively used to derive text summaries. The second paper, by Hahn and Reimer, presents an approach to text summarization based on knowledge representation structures derived from the TOPIC text understanding system. They have defined a set of salience operators, grounded in the semantics of a terminological logic, that are applied to knowledge databases produced by TOPIC in order to identify concepts, properties, and relationships playing a relevant role in the text. In the third paper, McKeown, Robin, and Kukich focus on linguistic summarization, the task of determining how to convey as much information as possible in a very short text. They describe the implementation of linguistic summarization in two systems: STREAK, for the basketball domain, and PLANDOC, for telephone network planning activities. The section concludes with a paper by Maybury

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 system. 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-