

## Book Review

KAREN SPARCK JONES AND PETER WILLETT (eds.)

*Readings in Information Retrieval*. Morgan Kaufmann Publishers, San Francisco, CA. 1997. ISBN 1-55860-454-5. £29.95. xv+589pp. Softbound.

In a 1983 letter to the members of the ACM SIG on Information Retrieval (IR), its chairman Gerard Salton (to the memory of whom the book here reviewed is aptly dedicated) expressed his concern that, while in the '60s "anyone concerned with the automatic processing of data and text seemed to be working in IR", this trend had seemingly reversed in favour of the database community, leaving only "a small number of active participants in IR work". Fifteen years later, although databases are still an important topic of research, there is little doubt that the excitement lies elsewhere, and (ironically enough) close to where it was in the '60s: if a poll were made, asking to name the computer science technology that has had the most profound impact on the general public in the last five years, Internet "search engines" would probably crush the opposition. Key contributors to the ever-growing success of the Internet, these children and grandchildren of the IR systems that Salton, Sparck Jones and other pioneers have been experimenting with since the '60s, have arisen an excitement that will no doubt contribute to the success of these *Readings*.

This authoritative work brings together a set of key papers that have contributed in laying the scientific foundations not only of Internet search engines, the conspicuous "tip of the iceberg" of nowadays IR, but also of the widely installed, if less glitzy, IR systems that are the essential everyday tools of the trade to many organisations. The choice of papers has favoured (a) technical papers that have had a determining influence on the field, rather than review papers; (b) papers proposing approaches and techniques that have since become established, rather than recent articles that have not yet withstood the test of time; (c) papers that altogether cover IR as a whole and, to a certain extent, "advanced" information management tasks such as document filtering, categorisation, and summarisation.

I especially appreciated the "categorisation scheme" that was chosen for this book. The papers are subdivided into eight chapters, respectively centred around the history of IR (Chapter 2), the key concepts of IR (Chapter 3), the evaluation of IR systems and techniques (Chapter 4), the formal models that underlie IR systems (Chapter 5), the techniques that are used for implementing them (Chapter 6), important experimental IR systems (Chapter 7), and recent extensions to the basic IR paradigm, including the above-mentioned "advanced" information management tasks, and the management of new types of documents, such as multimedia (Chapter 8). The final Chapter 9 contains a single, extremely thought-provoking paper discussing the inherent limits of IR. Of special importance, an index of no less than 15 pages is included at the end of the book (not a common feature of previously published article collections).

Each chapter is preceded by an introduction in which the editors present an overall view of the theme of the chapter, put the included articles in the context of this overall view, and provide key bibliographic references integrating those already contained in the articles themselves. These introductory chapters are no routine summaries, but offer to the reader an original synthesis and a stimulating interpretation of the theme of the chapter.

This is an important book, and one that was largely needed. Information retrieval is a disci-

pline that has matured slowly, although steadily, and one in which technological transfer, from research laboratories to the factory floor, has been slower than elsewhere. It is only recently that IR techniques or concepts (e.g. ranked retrieval) that had since long proven valuable in experimental settings, have been incorporated into commercial products. Because of this, many key papers that have determined *today's* developments, and are thus still essential reading, are scattered through rather old and sometimes hard to find journal volumes and conference proceedings (my estimate is that only a fraction of the key papers included here can be found in libraries with a less than robust information science and library science section). It is extremely appropriate, then, that a well-thought choice of important papers in this area should have been made available in compact form and at a reasonable price.

Unlike neighbouring disciplines (such as e.g. databases) for which a wide choice of textbooks is available, since the mid '80s IR has been suffering from a chronic lack of good, comprehensive, and up-to-date textbooks. Although not a textbook, these *Readings* will no doubt be an extremely useful support for lecturers offering IR courses, thanks not only to their balanced choice of material, but also to their excellent organisation into a natural thread that develops from the basic concepts to the recent, cutting-edge applications.

While this book will be definitely appreciated by practitioners of neighbouring disciplines (here I am especially thinking of databases, natural language processing, multimedia, hypertext, library and information science), by graduate students, and by information professionals, to whom it offers a unique and first-hand look at the development of this discipline, it is nothing less than an absolute must for anyone actively involved in IR, either in research or development.

FABRIZIO SEBASTIANI  
*Istituto di Elaborazione dell'Informazione*  
*Consiglio Nazionale delle Ricerche*  
*56126 Pisa, Italy*