ELECTRONIC PUBLISHING, VOL . 4(4), 183–184 (DECEMBER 1991)

Editorial

SPECIAL ISSUE ON INFORMATION RETRIEVAL

Information is continuing to grow exponentially and the increasing utilization of electronic and optical publishing technologies is making available large machine-readable document collections. There is a strong need for sophisticated and innovative retrieval systems which can provide satisfactory access to such amounts of stored information.

Information Retrieval (IR) has been developing from early bibliographic systems to the complex documentation systems found in current applications. At present, this is a very active research area, which encompasses a broad spectrum of topics, ranging from information retrieval theory, artificial intelligence, natural language processing, hypertext and multimedia, to a considerable variety of document storage and retrieval systems. Much progress has been achieved during the last decade on these and other related topics, feeding a rapidly growing market for information retrieval products and expertise.

The aim of this special issue of EP-odd is to focus on promising research efforts, system development and experimental results in the field. In the context of this issue, information retrieval is referred, primarily, to methods and techniques for searching relevant information out of large document collections. The issue consists of four papers that have been selected from the eight submitted and represents, in my opinion, a fair choice in order to show the breadth of research and development in this field.

In the first paper, "Nearest-neighbour searching in files of text signature using transputer networks", J. K. Cringean, R. England, G. A. Manson and P. Willett address the problem of improving the efficiency of the retrieval process by means of parallel computer architectures. They use a network of transputers, a high-performance microprocessor designed for parallel systems, to implement a two-stage search strategy. In the first stage, a screening text signature search is performed in order to prune the pattern-matching search carried out in the second stage. The prototype implementation is described along with experimental results discussing the increase in speed that can be obtained depending on different settings.

The second article, "NRT: News Retrieval Tool" is by M. Sanderson and Keith van Rijsbergen, who present the design and development of a retrieval system based on a client-server architecture and discuss users' experiences of interfacing to the Financial Times article database. It is interesting to see how new information retrieval techniques have been incorporated in this project, thus removing the well known limitations of conventional Boolean systems, and setting-up a system both powerful and easy to use.

In the third paper, "Issues of data modeling in information retrieval", M. Agosti, R. Colotti and G. Gradenigo propose a conceptual model of an information system. Essentially, a notion of schema is introduced, as opposed to specific document instances. The resulting two-level architecture can be browsed directly on the screen to get the required information or can be used to improve the query formulation process.

A fourth paper, logically part of this issue, will, for production reasons, appear in the next issue of EP-odd (volume 5, number 1). "Automatic structuring of text files" by

0894–3982/91/040183–02$06.00 *Received 2 December 1991*
© 1991 by John Wiley & Sons, Ltd.
© 1998 by University of Nottingham.

G. Salton, C. Buckley and J. Allen investigates the problem of providing flexible access to full-text databases in which the absence of homogeneous structure along with the great variety of information makes traditional techniques inadequate. In such situations, considerable improvements can be achieved by breaking down documents into individual text chunks, linking them together with appropriate relationships, and exploiting such relationships for both browsing and searching. With this goal, automatic tools are proposed for building a linked text structure from the existing source documents.

It is my hope that the articles in this special issue will show the reader that consistent improvements can be obtained by turning into practice new ideas from information retrieval research, and that there is still a large number of research issues that the reader may be motivated to pursue.

I would like to thank the contributors to this special issue for writing and revising their papers under severe time pressure. I also thank the reviewers for their prompt and constructive criticism, which helped to improve the content and the coherence of the entire issue. Finally, I would like to thank Rick Furuta for his encouragement, assistance and advice during the preparation of this special issue.

DARIO LUCARELLA, GUEST EDITOR