Editorial Message: Special Track on Information Access and Retrieval Systems

Fabio Crestani, Co-Chair
Dept. of Computer and Information Sciences
University of Strathclyde
Glasgow G1 1XH, UK

fabioc@cis.strath.ac.uk

Gabriella Pasi, Co-Chair
ITIM-CNR
via Ampère 56
20131 Milano, Italy

gabriella.pasi@itim.mi.cnr.it

1. NATURE OF THE TRACK

Currently, one of the most important and challenging problems in computer science is the definition of effective technologies, which support access to information. With the expansion of the Internet and the ever increasing dimension of the Web, effective tools for finding relevant information are urgently needed. This special track is concerned with theory and, in particular, applications of novel approaches to information access and retrieval.

Information access technologies, like for example Information Retrieval (IR) and Information Filtering (IF), aim at modelling, designing and implementing systems able to provide fast and effective content-based access to a large amount of information. Information can be of any kind: textual, visual, or auditory. The aim of such systems is to estimate the relevance of documents to a user's information need. This is a very hard and complex task, since it is pervaded with subjectivity, vagueness and uncertainty. Many existing systems offer a simple modelling of the information access task, which often privileges efficiency at the expenses of effectiveness. We think that a promising direction to improve information access effectiveness is to model the uncertainty and vagueness intrinsic in the process, and to make systems more adaptive, i.e. able to "learn" the users' concept of relevance. This special track is concerned with theory and, in particular, applications of novel approaches to information access and retrieval.

Currently there are two main venues where research and development in Information Access and Retrieval are presented: the ACM Conference on Research and Development in Information Retrieval (also known as SIGIR) and the annual Text Retrieval Conference (TREC) organised by the National Institute of Standards and Technology (NIST). These two conferences stand at the opposite ends of the spectrum of research and development in IR, dealing with mainly theoretical (SIGIR) or mainly implementation (TREC) aspects. In this track we try to present experiences that are more in the middle of this spectrum, where we believe many successful applications lie. Major topics of interest to this track include but are not limited to the following: models of information access and retrieval, applications of advanced information access and retrieval systems, multimedia and multimodal information access and retrieval, content-based information filtering, collaborative filtering, search engines, distributed information access and retrieval, information and data fusion, interfaces for information access, user modeling, and trials and best practice in information access and retrieval.

2. THE REVIEW PROCESS

We received 33 paper submissions to the special track. We did not expect so many papers and struggled to find 3 reviewers for each paper. Despite some problems with organizing and collecting more than 120 reviews, we managed to have each paper reviewed by at least 2 reviewers, with the large majority of papers receiving 3 blind reviews.

We believe our reviewer were rather tough, since we finally decided to accept only 15 papers, many of which subject to major revision. Such a number of accepted papers gave us an acceptance rate of 45%, well below the global average of past SAC tracks. This selective acceptance rate helped us assure the quality of the papers presented in the track. For this, we would like to thank the many reviewers (a list of which can be found in these proceedings) whose help was invaluable for selection process.

3. THE PAPERS

The special track comprises of 15 papers, divided in 4 sessions. In the first session, 4 papers dealing with document categorization and summarization are presented. The first paper, by J.M. Gomez Hidalgo, discusses some text categorization methods for filtering unsolicited

bulk Email. The second paper, by A. Becks, C. Seeling, and R. Minkenberg presents a comparative study of a document map based against a text-based access interface. A document map visualizes the vague semantic similarity structure of a corpus of documents. In the third paper, the authors D.L. Chan, R.W.P. Luk, W.K. Mak, H.V. Leong, E.K.S. Ho and Q. Lu propose the use of concept hierarchies, similar to Yahoo! categories, as a viable method to summarize multiple (related) documents for mobile clients. The fourth paper of the first session, by S.O. Sweeney, F. Crestani, and A. Tombros, presents the results of a study aimed at measuring the usefulness of presenting the results of an Information Retrieval Search on WAP mobile phones. The second session deals with problems related to Information Filtering, User Modeling, Information fusion. The first paper of this session, by M. Boughanem and M. Tmar, proposes an adapting filtering and learning model to simulate and evaluate an adaptive information filtering process. The second paper by H. Joho, Claire Coverson, M. Sanderson, and M. Beaulieu, deals with interactive query expansion; the authors propose a hierarchical presentation of expansion terms which are automatically generated from a set of retrieved documents. In the third paper, by P.J. Brown, G.J. F. Jones, deals with the problem of exploiting context information in Information Retrieval; the authors suggest new methods for improving the relevance of retrieved information in context-based retrieval. The fourth paper by H. Billhardt, D. Borrajo, and V. Maojo deals concerns the problem of data fusion; a system is presented which uses genetic algorithms to find a suboptimal combination of experts for a document collection. The third session collects papers dealing with Structured Document Retrieval and Data Structures. In the first paper, E. Kotsakis examines an XML document collection from the viewpoint of Information Retrieval, and attempts to adapt existing IR techniques to achieve more sophisticated search on XML documents. The second paper, by S. Ko and Y. Choy, proposes a new indexing method that supports attribute-based structured documents. The third paper, by G. Bordogna and G. Pasi, presents a flexible query language for expressing soft selection conditions on documents structured in logical sections. The fourth paper by S. Watanabe, T. Miura deals with the definition of data structures: the authors present a reordering technique that improves sequential processing to B-tree files. The fourth session collects papers on Image Retrieval and Spelling Correction. The first paper by M.A. Nascimento and V. Chitkara proposes two variations of a new image abstraction technique based on signature bit-strings and an appropriate similarity metric for color-based image retrieval. In the second paper by D. Stan and I.K. Sethi a new type of image feature is proposed, which consists of patterns of colors and intensities that capture the latent associations among images and primitive features in such a way that the noise and redundancy are eliminated. In the third paper, by P. Ruch, the author compares different strategies for correcting spelling errors; a system is presented, which has been developed to take advantage of the entire context surrounding mispelling.

4. THE TRACK CHAIRS

Dr. Fabio Crestani is a Reader at the Department of Computer and Information Science of the University of Strathclyde in Glasgow, Scotland. Before joining Strathclyde University he held research fellowship positions at the Rutherford Appleton Laboratory of the CLRC (UK), the International Computer Science Institute in Berkeley (USA), and the University of Glasgow (UK). From 1992 to 1997, he was Assistant Professor at the University of Padova (Italy). He holds a degree in Statistics from the University of Padova, and a MSc and PhD in Computing Science from the University of Glasgow.

Fabio Crestani has co-edited 3 books and published over 60 refereed publications in the areas of information retrieval, hypermedia, and digital libraries. He has collaborated with a number of National and International research projects and has a number of academic and industrial contacts in Europe and in USA. He has also been a member of the organizing and program committee of several conferences and workshops. More up-to-date information on his research interests, including online copies of his publications, can be found at the URL: http://www.cs.strath.ac.uk/~fabioc/.

Dr. Gabriella Pasi is a researcher at the Institute for Multimedia Information Technologies of the National Research Council in Milano, Italy, since 1985. She holds a degree in Computer Science from the University of Milano, Italy, and a PhD in Computer Science from the University of Rennes, France. She has been a contract professor at the University of Trento, Italy. She is a member of the executive board of the Information Technologies Group of the Italian Electrotechnical Association (AEI). She is also a member of the board of the EUSFLAT Association (European Association of Fuzzy Logic and Technology), and a member of the North American Fuzzy Information Processing Society. She is a member of program committees and of organizing committees of several international conferences.

Gabriella Pasi has co-edited 4 books, one special issue and published over 80 refereed papers in international journals, in books and in proceedings of international conferences. She is leading projects aimed at the modelization of uncertainty and imprecision in systems for Information Retrieval, and for the management of Object Oriented DataBases. Her current research interests include modeling of flexible information retrieval systems and relevance feedback mechanisms, management of imprecision and uncertainty in databases, information fusion, fuzzy logic, multi criteria decision making, e-commerce. Additional information on Gabriella Pasi's research can be found at the URL: http://www.itim.mi.cnr.it/Staff/Pasi/home.html.