Report on the 24th European Colloquium on Information Retrieval Research (ECIR 2002)

Fabio Crestani
Department of Computer and Information Sciences
University of Strathclyde
Glasgow, Scotland, UK
fabioc@cs.strath.ac.uk

Mark Girolami
School of Information and Communication Technologies
University of Paisley
Paisley, Scotland, UK
mark.girolami@paisley.ac.uk

Keith van Rijsbergen
Department of Computing Science
University of Glasgow
Glasgow, Scotland, UK
keith@dcs.gla.ac.uk

1 Introduction

Information retrieval (IR) is the science and technology concerned with the effective and efficient retrieval of information for the subsequent use by interested parties. The central problem in IR is the quest to find the set of relevant documents, amongst a large collection, containing the information sought thereby satisfying an information need usually expressed by a user with a query. The documents may be objects bearing information in any medium, text, image, audio, or, indeed a mixture of all three.

The annual colloquium on IR research has provided for 24 years an opportunity for both new and established researchers to present papers describing work in progress or final results. These Colloquia were established by the BCS IRSG (British Computer Society Information Retrieval Specialist Group), and named the Annual Colloquium on Information Retrieval Research. Recently, the location of the colloquium has alternated between the United Kingdom and continental Europe to reflect a growing interest in research collaboration beyond the United Kingdom boundaries. In order to reflect also in the name this growing European orientation of the event, the Colloquium was renamed "European Colloquium on Information Retrieval Research" (ECIR) from 2001.

ECIR 2002 was held in Glasgow on March 25-27, 2002. Since the inception of the colloquium in 1979 the event has been hosted in the city of Glasgow on four separate occasions. However, this was the first time that the organization of the colloquium had been jointly undertaken by three separate computer and information science departments of three of the four universities present in the Greater Glasgow area: the University of Glasgow, the University of Strathclyde, and the University of Paisley. This is an indication of the collaborative nature and diversity of IR research within the universities of the West of Scotland, which by hosting three growing IR research groups is now one of the *power houses* of IR research in Europe.

The organizers of ECIR 2002 saw a sharp increase in the number of good quality submissions in answer to the call for papers over previous years and as such 52 submitted papers were each allocated three members of the program committee for double blind review of the manuscripts. A total of 23 papers were eventually selected for oral presentation at the colloquium in Glasgow which gave an acceptance rate of less than 45% and ensured a very high standard of the papers presented. These papers have been included in full in the proceedings of the event, edited by the event organisers: Fabio Crestani, Mark Girolami, and Keith van Rijsbergen, and published in a Springer-Verlag Lecture Notes in Computer Science Series volume. The proceedings are also available online through the Springer-Verlag Web site at `http://http://link.springer.de/link/service/series/0558/tocs/t2291.htm`. Browsing through the table of contents the reader will note that in addition to contributions from European authors just over 20% of the papers presented were authored by researchers from the USA and Asia which added a more international flavor to ECIR 2002.

In this report we briefly highlight the thematic areas covered in the colloquium and summarise some of the issues discussed during the sessions.

2 The Sessions

The colloquium was organised in 9 sessions comprising between 2-3 papers each. This clustering of papers in small sessions enabled the audience to focus on very specific areas of IR research. The sessions dealt with: Multimedia IR, Web IR, Query Modification, Soft Computing and IR, IR models, Document Categorization, Structured Document Retrieval, Cross Lingual IR, and Interactive IR Systems.

The colloquium commenced with an *invited talk* by Rik Belew from the University of California at San Diego. The choice of the invited speaker was motivated by many and different reasons, not least the popularity of his recent book on IR. In "Finding out about: search engine technology from a cognitive perspective" Belew takes an original perspective in the explanation of how to build tools for searching collections of text and other media. In the process of explaining how to design and implement these tools, he asks us to give a close look at the properties of textual documents seen as single elements and as part of large collections. He shows that the construction of effective IR systems requires not only knowledge of the statistical and mathematical properties of the linguistic phenomena, but also an appreciation of the cognitive foundations we bring to the task as language users.

The title of the invited talk delivered by Rik Belew was "Restoring the context between writer and reader" and it provided a broad and novel perspective on some of the important issues which require to be considered in current and future directions of IR research. He then went on to speak about hierarchies and how they are a natural way for people to organize information albeit that different people and organizations tend to construct different conceptual hierarchies (e.g., contrast Yahoo!, UseNet news and OpenDirectory hierarchies). He pointed out that whilst there are often significant commonalities it is in general quite difficult to fully reconcile them. The problem of embedding a narrow and focused topical hierarchy into a broader one was considered, and two algorithms for accomplishing this task were discussed.

The colloquium started in earnest after the invited talk. The first session was devoted to *Multimedia IR* with the presented papers focusing on indexing and retrieval of images and melody extraction. Work in this area is at the forefront of IR and Digital Libraries research and the papers in this session provided some insights into progress issues such as integration of textual and content-based image retrieval for searching partially indexed image databases as well as the impact on retrieval performance of graphical sketches when combining varying shape representations. The session devoted to *Web Based Retrieval* had contributions which considered the combination of both content and link structure employing the Bayesian inference network formalism, an improved method for computing the PageRank algorithm, and dynamic webpage content customisation. There was some rigorous discussion during the *Query Modification* and *Soft Computing* sessions where papers on profile reformulation, the use of implicit feedback, query optimisation employing genetic algorithms, and a connectionist based adaptive IR model were presented. The sessions dedicated to *IR Models* and *Categorisation* had a strong bias towards probabilistic representation and modeling. The sessions covered Bayesian network models, both flat and hierarchic mixture models, uncertainty based term selection, and term frequency normalisation employing the Pareto distribution. The remaining sessions on *Structured Documents*, *Cross-Language* and *Interactive Systems* each had two presentations which provided the colloquium participants the opportunity to discuss at some length both the experimental and theoretical facets of the papers during the question time.

3 Social Programme and Awards

The colloquium was not only a serious research event, it was also a joyful occasion for the more than 90 participants to chat in friendly surroundings about their work, life, and latest developments. First of all, the colloquium venue was quite central in creating this feeling of community. ECIR 2002 was held at the Teachers Building, a beautiful listed building with a relatively small and cosy lecture theatre right in the center of Glasgow. This was supported by a well planned social programme, which started with a visit and reception in "grandeur" of the Glasgow City Chambers on the first day, followed by a dinner in the impressive Corinthian, on the second day.

Two prizes were awarded during the dinner. The best student paper prize was awarded to Victor Lavrenko of the University of Massachusetts at Amherst (USA) for his paper on optimal mixture models. The best student award was sponsored by Memex Technology, a local (East Kilbride) IR company with a strong interest in IR research and young IR talents. The best paper award went to Eero Surmonen of the University of Tampere (Finland), for his paper on a retrospective evaluation method for exact-match and best-match queries applying an interactive query performance analyser. The best paper award was sponsored by Microsoft Research, Cambridge (UK).

4 Conclusions

We, the organisers, are proud of the success of ECIR 2002. We believe we have succeeded in keeping the quality of the presentations and of the discussions at the highest levels. We are confident that the importance of ECIR as a research forum will grow in importance beyond the boundaries of Europe in the future years and we believe it will soon be recognised as a truly international event on a par with the few other international IR conferences.

Acknowledgements

The organisers would like to thank all the authors who submitted their work for consideration and the participants of ECIR 2002 for making the event a great success. Special thanks are due to the members of the program committee who worked so hard to ensure the timely review of all the submitted manuscripts. We also thank the sponsors CEPIS-IR (Special Interest Network on Information Retrieval of the Council of European Professional Informatics Societies), the British Computer Society IR Special Interest Group, Memex Technology, and Microsoft for their generous financial support of the colloquium, and Glasgow City Council for civic hospitality.

Finally thanks are due to the local team of student volunteers, secretaries and information officers whose efforts ensured the smooth organization and running of the colloquium.