Report on the
25th European Conference on Information Retrieval Research (ECIR-03)

Fabrizio Sebastiani
Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche
Pisa, IT
E-mail: `fabrizio.sebastiani@isti.cnr.it`

1 Introduction

The 25th European Conference on Information Retrieval Research (ECIR-03 – http://ecir03.isti.cnr.it/) took place in Pisa, Italy, on 14-16 April, 2003, and was organized by the Italian National Council of Research.

The ECIR series was initially established by the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG) under the name "Annual Colloquium on Information Retrieval Research", and always held in the United Kingdom until 1997. Since 1998 the location of the Colloquium has alternated between the United Kingdom and the rest of Europe, in order to reflect the growing European orientation of the event. For the same reason, in 2001 the event was renamed "European Annual Colloquium on Information Retrieval Research". Since 2002, the proceedings of the Colloquium are being published by Springer Verlag in their Lecture Notes in Computer Science series ([1, 2]). As of 2003 BCS-IRSG has decided to rename the event "European Conference on Information Retrieval Research", in order to reflect what the event had slowly turned into, i.e. a full-blown conference with a European program committee, strong peer reviewing, and a (mostly) European audience.

ECIR-03, the 25th "Silver Jubilee" edition, represented a record-breaking success. The number of submitted papers was the highest in the series (101), which resulted also in a record-low acceptance rate (31% for full papers), thus ensuring a high qualitative standard. The number of participants was also the highest in the series (120), notwithstanding two major international crises (war in Iraq and SARS) which were rampant at the time and which somehow discouraged attendance from non-Europeans.

ECIRs have a tradition of being strongly student-oriented events, i.e. settings in which students and young researchers have the scene mostly for themselves and feel (or should feel) unintimidated in discussing their ideas with more experienced researchers; many IR researchers from the UK or neighbouring areas have indeed discussed their first ideas at the Colloquium, and have since gone on to give major contributions to IR research. In keeping with the tradition, students were very well represented in ECIR-03, since 20 out of the 31 full papers and 13 out of the 16 short papers had (and mostly were presented by) a full-time student as the main author. The traditional student focus of ECIR was also enforced through (i) below-cost registration fees for students; (ii) a rich student grant program, which allowed 26 European students (more than 20% of the total attendance of the conference) to be supported by a grant covering registration fee, lunches, banquet, proceedings, and pocket money; and (iii) through the presentation of the Best Student Paper Award (in keeping with its student focus, ECIR-03 did not have a Best Paper Award). This award, sponsored by Scirus (Elsevier group), was presented to Henrik Nottelmann (University

of Duisburg-Essen, DE) for his paper "From uncertain inference to probability of relevance for advanced IR applications", co-authored by Henrik's supervisor Norbert Fuhr.

2   The keynote talks

In order to mark the Silver Jubilee of ECIR, Karen Spärck Jones (University of Cambridge, UK) gave an initial keynote presentation (titled "Document retrieval: shallow data, deep theories; historical reflections, potential directions") in which she presented an historical perspective on the development of statistically based retrieval in the last 40 years and more. Spärck Jones especially looked at the consequences of two ideas that, trivial as they make look today, were revolutionary at the time they were proposed, i.e. (i) that the words occurring in the documents should be used for representing documents and querying them, and (ii) that the frequency of occurrence of these words and the distribution of these occurrences in the documents and in the collection are of key importance for detecting relevant documents. Spärck Jones discussed how these two ideas not only marked the birth of fully automated retrieval (as opposed to the once-current brand of retrieval that required manual indexing of documents) but also formed the basis for most of the later developments in our field, including the most recent ones in Web retrieval and language modelling.

The second keynote talk (titled "Annotation and Retrieval of Structured Video Documents") was given by Alberto del Bimbo (University of Firenze, IT), who gave a broad presentation on his work on video annotation for supporting retrieval by content of video segments, and illustrated this work by drawing examples from a specific application he has worked on, i.e. the detection and recognition of soccer highlights. Del Bimbo presented an approach to fully automatic annotation of soccer games videos, based on the automatic recognition of playfield zones and on the detection and classification of other low-level features such as camera motion parameters. He demonstrated that the resulting system displays an impressive efficiency and effectiveness in addressing mid and high-level queries about required highlights, thus making it viable for applications (such as the live broadcasting of sports events) in which editors need to rapidly and effectively access highlights of video material that has just been produced.

The full text of both keynote talks is included in the published Proceedings [2].

3   The papers

A total of 31 full papers and 16 posters were presented at the conference. Collectively, they are indicative of the wide range of issues being tackled in current IR research, and include both theoretical and experimental work in several media (text, hypertext, structured text, multi-lingual text, spoken text, images, music) and in several tasks (search, retrieval, clustering, categorization, both content-based and collaborative filtering, summarization, information extraction, question answering, topic detection and tracking, visualization), either in centralized or in distributed environments, and tackling either effectiveness or efficiency issues.

Unsurprisingly, the area with the highest number of papers was that of machine learning (ML) for IR; this confirms a trend which has also been observed at SIGIR conferences, and which has forced the SIGIR-04 Program Chairs to nominate no less than four area coordinators (or metareviewers) for this topic. Within the application of ML to IR, the one that has drawn the largest attention has been text categorization (TC – the only topic which had two sessions of its own at ECIR-03); among others, approaches to TC based on Kullback-Leibler divergence, language modelling extensions to Naive Bayes methods, entropy-based discretization of continuous attributes, or active learning, have been presented. Other subfields of ML&IR were also objects of presentation; these included collaborative filtering, recommendation technology as applied to on-line store design, clustering for multi-document summarization, information extraction from spoken documents, and others.

One session was also devoted to formal models and language models of IR, a topic which

in Europe has traditionally enjoyed a special popularity, especially on the formal models side; presentations discussed large scale evaluations of propositional logical models, new approaches to drawing uncertain inference for IR modelling, and specialized representations of spatio-temporal information for topic detection and tracking.

A topic which also enjoys a strong popularity in Europe is user studies for IR; one session was devoted to it, featuring papers on tilebar-like representations for supporting within-document retrieval, user studies of relevance feedback for image retrieval applications, and tools for the graphical visualization of Web search results.

A session was also devoted to the applications of natural language engineering to IR; contributions included analyses of the impact of various types of phrase indexing on text categorization, comparative evaluations of stemming and decompounding methods for collections of long documents, analyses of the importance of term proximity for keyword-based search, and linguistic analysis for question answering applications.

Two topics which have recently been quite popular in the IR community, i.e. Web search and structured document retrieval, were devoted one session each. The former featured a paper on hierarchical classification of Web pages and two related papers discussing the usefulness of links for Web search and Web classification, a theme which has aroused a lot of interest recently in view of the rather uninspiring results obtained in the TREC Web track by systems making use of link-based information. The latter session included papers on new methods for the retrieval of structured documents, user studies concerning the structured document retrieval task, and experiences on building test collections for this task.

Finally, a session was devoted to architectural and efficiency issues for IR, and dealt with compression issues for text collections, and data structures in secondary storage for supporting searches in Web directories or for supporting approximate string processing.

Some topics of traditional IR interest were under-represented in the ECIR-03 program. For instance, while text was the focus of attention of the majority of papers, few papers tackled issues in multimedia IR. Since the distribution of the papers across the topics largely reflected the distribution of the *submitted* papers, this may indicate a smaller degree of activity on these topics on the part of European researchers with respect to their non-European colleagues, or a tendency from the relevant researchers to submit to other more specialized conferences. Also, while traditional tasks in text management (e.g. text categorization) had a strong presence, tasks of the cutting-edge brand such as question answering, topic detection and tracking, and summarization, had a smaller presence than they have e.g. at recent SIGIR conferences. We all hope that this fact can stimulate European funding agencies to support European research in these areas more strongly.

Concerning the geographical distribution of authorship, 77% of the papers indeed came from Europe, with the United Kingdom (20%), Italy (14%), Spain (11%), and Germany (6%), being the countries with the highest representation (percentages are calculated by the affiliation of each co-author). Among the non European papers, Asia was (somehow surprisingly) the most highly represented region (14%), while other regions had a smaller representation (4% from North America, 3% from the Middle East, and 1% from South America). Overall, ECIR-03 authors came from a total of 22 different countries across the world.

4 The panel

The ECIR-03 program also included a panel on "The future of IR research in a European context", coordinated by Keith van Rijsbergen (University of Glasgow, UK), and with Maristella Agosti (University of Padova, IT), Ayse Göker (Robert Gordon University, UK), Cornelis H.A. Koster (University of Nijmegen, NL), Peter Ingwersen (Royal School of Library and Information Science, DK), and Alan Smeaton (Dublin City University, IE). The panel tackled the role of Europe in the broader context of IR research, especially in the light of the current funding policy of the Commission of the European Communities, whcih seems to have largely diverted its attention from solid experimental IR in favour of other emerging fields such as e.g. the "Semantic Web".

The discussion that followed was very lively, and touched on issues such as the nature and future of the ECIR conference series and of the ESSIR (European Summer School on Information Retrieval) series, and the takeup of IR research results by European companies. A detailed report on this panel will appear in an upcoming issue of the *Informer*, the on-line newsletter of the BCS-IRSG (http://irsg.eu.org/).

The proceedings of ECIR-03 [2] are available from Springer Verlag at the price of 70 Euro. A special issue of the Information Retrieval journal (Kluwer Academic Publishers) devoted to ECIR-03 is also forthcoming in early 2004, which will contain substantially revised and extended versions of seven among the papers that were ranked higher by the ECIR-03 Program Committee.

The next edition of ECIR (http://ecir04.sunderland.ac.uk/) will take place at the University of Sunderland, UK, from April 5 to 7, 2004, and will be organized by Sharon McDonald and John Tait.

References

[1] Fabio Crestani, Mark Girolami, and Cornelis J. van Rijsbergen (eds.). *Proceedings of the 24th European Colloquium on Information Retrieval Research*, Glasgow, UK, 2002. Published by Springer Verlag, Heidelberg, DE, as volume n. 2291 in the *Lecture Notes in Computer Science* series.

[2] Fabrizio Sebastiani (ed.). *Proceedings of the 25th European Conference on Information Retrieval Research*, Pisa, IT, 2003. Published by Springer Verlag, Heidelberg, DE, as volume n. 2633 in the *Lecture Notes in Computer Science* series.