Report on the 27th European Conference on Information Retrieval Research (ECIR 2005)

David E. Losada
Grupo de Sistemas Inteligentes
Departamento de Electrónica y Computación
Universidad de Santiago de Compostela
Spain
dlosada@dec.usc.es

Juan M. Fernández-Luna
Grupo de Tratamiento de la Incertidumbre
en Inteligencia Artificial
Departamento de Ciencias de la Computación
e Inteligencia Artificial
Universidad de Granada
Spain
jmfluna@decsai.ugr.es

1 Introduction

The 27th European Conference on Information Retrieval Research (ECIR 2005), celebrated in Santiago de Compostela (Spain) on 21-23 March 2005, was jointly organized by the University of Santiago de Compostela (Spain) and the University of Granada (Spain), under the auspices of the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG). The conference was co-chaired by David E. Losada (University of Santiago de Compostela) and Juan M. Fernández-Luna (University of Granada) and was celebrated in the Technical School of Engineering of the University of Santiago de Compostela.

This conference 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". The colloquium was held in the United Kingdom each year until 1998, when the event was organised in Grenoble, France. Since then the conference venue has alternated between the United Kingdom and Continental Europe, reflecting the growing European orientation of ECIR. For the same reason, in 2001 the event was renamed "European Conference on Information Retrieval Research". In recent years, ECIR has continued to grow and has become the major European forum for the discussion of research in the field of Information Retrieval.

In terms of submissions, ECIR 2005 has been a record-breaking success, since 124 full papers were submitted in response to the call for papers. This is the highest number of submissions received in the ECIR series. ECIR 2005 established also a call for posters and there were 41 poster submitted. Paper and poster submissions were received from across Europe and further afield including North America, South America, Asia and Australia, which is a clear indication of the growing popularity and reputation of the conference. All papers and posters were reviewed by at least three reviewers. Out of the 124 submitted papers, 34 (27%) were accepted. Likewise, 17 (41%) posters were accepted.

ECIRs use to be strongly student-oriented events. The participation of students in the conference is usually encouraged through low-cost registration fees for students, a student grant program and the presentation of the Best Student Paper Award. Students were very well represented in the ECIR 2005 program, since 22 out of 34 full papers and 10 out of 17 posters involve a full-time student as the primary author, which means that the traditional student focus of the conference has been very well preserved.

Following the tradition of the most recent ECIRs, ECIR 2005 proceedings were published by Springer in their Lecture Notes in Computer Science series [2].

The conference was sponsored by the Information Retrieval Specialist Group of the British Computer Society (BCS-IRSG), the University of Granada, the Council of European Professional Informatics Societies (CEPIS), Microsoft Research, Sharp Laboratories of Europe Ltd. and the European Research Consortium for Informatics and Mathematics (ERCIM).

2 Keynote Speakers

The conference scientific program was opened by Keith van Rijsbergen (University of Glasgow, UK), who presented a keynote speech entitled "A probabilistic logic for information retrieval". Keith showed how logic, probability and geometry can be combined to generate a formal framework for specifying information retrieval models. More specifically, van Rijsbergen discussed how a vector space can be looked at in terms of states and observables leading to a formal framework whose interpretation of relevance and aboutness deviates from the traditional one. More details on this research can be found in Keith's keynote paper in the proceedings [4] and exhaustive discussions on this subject are also available in Keith's recent book [3].

The second keynote speech was given by Ricardo Baeza-Yates, who discussed several new techniques to improve the web search process based on the analysis of query logs. Server logs of search engines store traces of queries submitted by users (including queries themselves and web pages selected in their answers). On the other hand, website logs store queries and later actions (obtained from search engines referrers or from internal search boxes). Ricardo presented two methods to apply this logged information for analyzing and clustering queries. The first one leads to changes to improve the text and structure of a website and the second one does relevance ranking boosting and query recommendation in search engines. Further details are available in [1].

3 Full papers

A total of 34 papers were presented at the conference. The topics covered by papers both theoretical and experimental work in several IR tasks (summarization, classification, fusion, retrieval, etc.) and in different media (text, images, video, hypertext, structured text, etc.). Papers proposed research efforts either in centralized or distributed environments and dealing with efficiency or effectiveness issues.

The first session of the conference was dedicated to peer-to-peer systems. This research topic has recently become quite popular in the IR community. The first paper proposed an history-based P2P search algorithm and topology adaptation mechanism for overcoming the scalability problem of these systems. The second paper presented a suite of testbed for evaluating P2P systems and the last paper of the session dealt with text-based federated search of digital libraries using relevance-based ranking.

A session was dedicated to text summarization and contributions included proposals for profile-biased multi-document summarization, summarization based on word clusters and ranking algorithms, headline generation systems and a summarization system that extracts textual descriptions from Medline records.

A topic which also enjoys a strong popularity is text classification. In ECIR 2005 a session was dedicated to the topics of text classification and fusion. This session contained papers about supervised classification on document collections containing also junk documents, compression-based classification and the effect of correlation on data fusion for multiple retrieval results.

User studies and evaluation were devoted one session. One paper presented a user study to determine the usability of an ontology as a search tool. Other paper was dedicated to analyze query similarity in collaborative web search and the other two contributions dealt with a probabilistic interpretation of precision, recall and F-score and approaches to human evaluation of search engine relevance.

Multimedia retrieval was also the topic of another session, including papers on parsing football video structures, fractional distance measures for content-based image retrieval and combination of visual semantics and texture characterizations for image retrieval.

Another session was dedicated to web retrieval and its contributions included applications of associative relationships on the clickthrough data to improve web search, studies on factors affecting web page similarity and the utilization of phrase and proximity terms for boosting web retrieval.

Four sessions included heterogeneous proposals in the context of different IR models and techniques. The contributions presented in these sessions covered aspects such as formal studies on the web structure, probabilistic approaches to schema matching, encoding XML in vector spaces, formal models for structured document retrieval, logic-based systems for retrieval and question answering, latent semantic indexing models for efficient search, relevance feedback on small displays, novel term frequency normalizations, indexing with document identifier reassignment based on dimensionality reduction, scalability influence on retrieval models, muti-word units in interactive IR, dictionary-based cross-language IR and the application of key terms in top retrieved documents to improve retrieval effectiveness.

Paper submissions were received from across Europe and further afield including North America, South America, Asia and Australia, which is a clear indication of the growing popularity and reputation of the conference. More specifically, 76% of the papers came from Europe, with the United Kingdom (23%), France (14%), Germany (7%), Finland (5%), Spain (5%) and Ireland (5%), being the countries with the highest representation in the scientific program. Among non European papers, the highest presence comes from the United States (14%). The ECIR program contains also papers from Singapore, Canada and China.

It is also remarkable the increasing presence of research papers from leading companies.

Thanks to the generosity of the Council of European Professional Informatics Societies (CEPIS), we were able to offer travel grants for supporting the participation of 10 students in ECIR 2005. The ECIR 2005 Best Student Paper Award was presented to Gilad Mishne, from the University of Amsterdam, for his paper "Boosting Web Retrieval through Query Operations". This paper was co-authored by Gilad's supervisor, Maarten de Rijke.

4 Posters

The poster submission process was also very competitive (acceptance ratio: 41%). A total of 17 posters were presented at the conference. The ECIR 2005 poster program had a strong emphasis on web and multimedia retrieval and conference attendants could discuss late breaking results on terabyte IR platforms, search engines, distributed IR systems, video and image retrieval, etc.

Concerning the geographical distribution of poster authorship, 76% of the posters came from Europe (most from Ireland 23%, UK 16%, The Netherlands 11%, Spain 7%). Non-european posters came from China (11%), Japan (5%) and Singapore (5%).

References

[1] R. Baeza Yates. Applications of web query mining. In D.E. Losada and J.M. Fernández-Luna, editors, *Advances in Information Retrieval. Proc. of the 27th European Conference on Information Retrieval Research*, number 3408 in Lecture Notes in Computer Science. Springer Verlag, 2005.

[2] D.E. Losada and J.M. Fernández-Luna, editors. *Advances in Information Retrieval. Proc. of the 27th European Conference on Information Retrieval Research.* Number 3408 in Lecture Notes in Computer Science. Springer Verlag, 2005.

[3] C.J. van Rijsbergen. *The Geometry of Information Retrieval.* Cambridge University Press, 2004.

[4] C.J. van Rijsbergen. A probabilistic logic for information retrieval. In D.E. Losada and J.M. Fernández-Luna, editors, *Advances in Information Retrieval. Proc. of the 27th European Conference on Information Retrieval Research*, number 3408 in Lecture Notes in Computer Science. Springer Verlag, 2005.