REPORT ON THE FIRST 5 YEARS OF THE TRACK ON INFORMATION ACCESS AND RETRIEVAL OF THE ACM SYMPOSIUM ON APPLIED COMPUTING

Fabio Crestani° and Gabriella Pasi*
°Department of Computer and Information Sciences
University of Strathclyde, Glasgow, UK
f.crestani@cis.strath.ac.uk

* DISCO, Università degli Studi di Milano Bicocca
Via Bicocca degli Arcimboldi 8, Milano, Italy
pasi@disco.unimib.it

Abstract

We briefly report on 5 years of successfully running the ACM Symposium on Applied Computing (SAC) track on Information Access and Retrieval (IAR). We believe the track plays a valuable role as a forum for the presentation of research and development work in Information Retrieval. The growing number of submissions indicates that there is a need for such forum.

1 A Brief History of the SAC-IAR Track

Readers of SIGIR FORUM know very well that Information Retrieval (IR) is a very hard and complex task and that a large volume of research has attempted to explain and tackle. Nowadays, research in IR is central to the design and development of advanced information access technologies and spans a number of research topics including document modelling, document classification and categorization, system architecture, user interfaces, data visualisation, languages, topic detection, evaluation, etc.

In the past 30 years, IR has grown well beyond its primary goals of indexing and searching textual documents in static bibliographic collections, and has moved away from the perception of being the narrow area of interest of librarians and information experts. Information access technologies, and in particular IR, are currently being used in many different application contexts that go far beyond the initial scope of their design. Moreover, the development of the Internet has made urgent the problem of designing systems that effectively retrieve from the WWW information that is relevant to users' needs. The application of models and techniques proposed and tested in standard experimental contexts to new application areas is a very challenging task that we believe is worth of great attention by the researchers.

Currently there are two main venues where the research and development in IR is presented: the ACM Conference on Research and Development in Information Retrieval (also known as SIGIR) and

the annual Text Retrieval Conference (TREC for short) organised by the National Institute of Standards and Technology. These two conferences stand at the opposite ends of the spectrum of research and development in IR, dealing with mainly theoretical (SIGIR) or mainly evaluation related (TREC) aspects. The special track on Information Access and Retrieval (IAR) of the ACM Symposium on Applied Computing (SAC) lies, instead, in the middle of this spectrum, where we believe very valuable and successful applications of IR research lie. These applications might sometime not be theoretically very advanced and their effectiveness and/or efficiency might be sometime relatively unremarkable when measured by TREC standards, but their contributions to research and development in IR is still very valuable and useful. In fact, lessons learned from these applications can be very useful not only to practitioners, but also to researchers, as a testing ground for new models and methods that are still too weak to be presented at SIGIR or experimented with at TREC.

The IAR Track was firstly organized at the ACM SAC 2002, and since then it has been running at SAC every year. While papers in all aspects of research and development of IR are considered, it is concerned mainly with the theory, implementation and evaluation of IR to novel application areas and novel contexts of information access. Its main aim is to provide a forum for allowing researchers and practitioners in the IR field to present their experiences in the middle of the theory-practice spectrum of IR research and development. Since 2002 papers accepted for presentation and publication in the proceedings published by ACM and included in the ACM Digital Libraries have concerned several important and hot areas of research and development in IR and closely related areas, including multimedia, information filtering, document summarisation, information extraction, effectiveness evaluation, classification, web searching, and cross language IR. Since 2002, the track has always been co-chaired by the two original proposers of the track: Fabio Crestani and Gabriella Pasi.

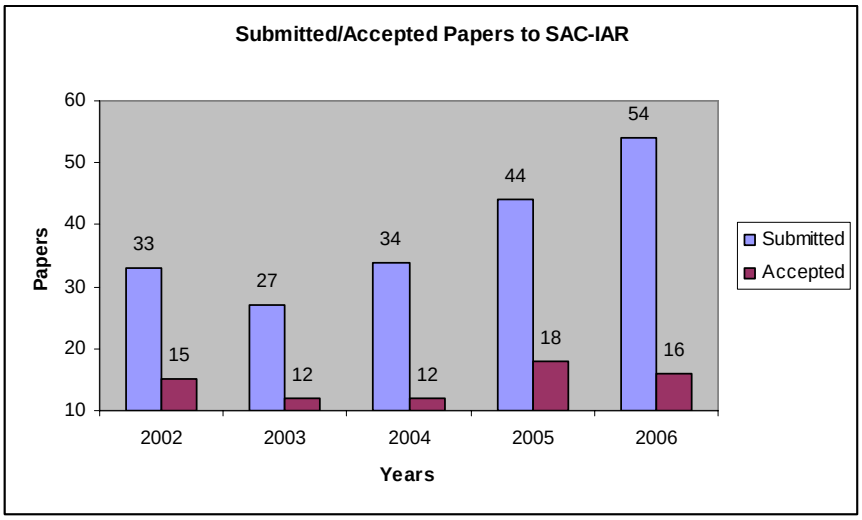

Figure 1

The graph in Figure 1 shows the number of papers submitted to the IAR track since 2002. As it can be seen, the number of submission has been steadily increasing (with the exception of 2003, when a short decline in submission could be explained by many factors external to the track, such as SARS, for example). The high and increasing number of submissions over the last 5 years is also proof of the importance of the IR research field and in particular of its applications. Conversely, the number of

papers accepted and the consequent acceptance rate has been almost constant, thus enabling us to concentrate more and more on accepting papers of very high quality. For this, our gratitude goes to the over 50 member of the track Program Committee whose help was invaluable for the selection process. In fact, every paper was reviewed by at least 3 members of the Program Committee and only papers with at least two positive reviews were considered of acceptable quality to be considered for acceptance.

Finally, within the context of SAC the IAR track has been growing in prestige and importance, accounting for a sizeable portion of all papers submitted and accepted, also in relation to other long established and running track.

2 SAC-IAR 2006

The 2006 edition of the Symposium on Applied Computing track on Information Access and Retrieval (SAC-IAR 2006) was held in Dijon, France, on 23-27 April. So far, it has been the most successful edition of the track, attracting 54 paper submissions and confirming the trend of increasing number of submissions since the start of this initiative. Only 16 papers were accepted, thus giving the track an acceptance rate of less than 30%, the lowest we ever had in previous tracks. Given the high quality of the submission and the recent introduction of posters at SAC, we also accepted 5 posters. As in many other ACM conferences, including SIGIR, each poster is allowed two pages in the proceedings.

The 16 papers were presented in 4 sessions of 4 papers each. Each paper had 20 minutes for presentation and 5 minutes for questions and discussion. The questions and discussion often went well beyond the 5 minutes allocated, showing how interested the audience was in the work presented. We do not attempt to describe the content of these papers in this report. We simply outline the topics dealt in each session.

The first session comprised papers that attempt to use the document or the query structure in a number of tasks, ranging from document summarisation, discovery of unexpected information, query transformation and retrieval effectiveness. The second session presented papers tackling information classification, categorisation and filtering for email, novelty detection and P2P file sharing. While these two sessions were rather homogenous in topics, the remaining sessions were more heterogeneous. Session 3 presented three papers dealing with web searching, addressing topics like web search ranking, web pages template detection, and clustering of web snippets. This section also presented a paper dealing with light stemming of a number of different languages for cross-language IR. Session 4 presented three papers on multimedia IR, dealing with retrieval and filtering of music, logo and trademark images, and images of illuminated manuscripts. This session also included a paper on a translation technique for out-of-vocabulary words.

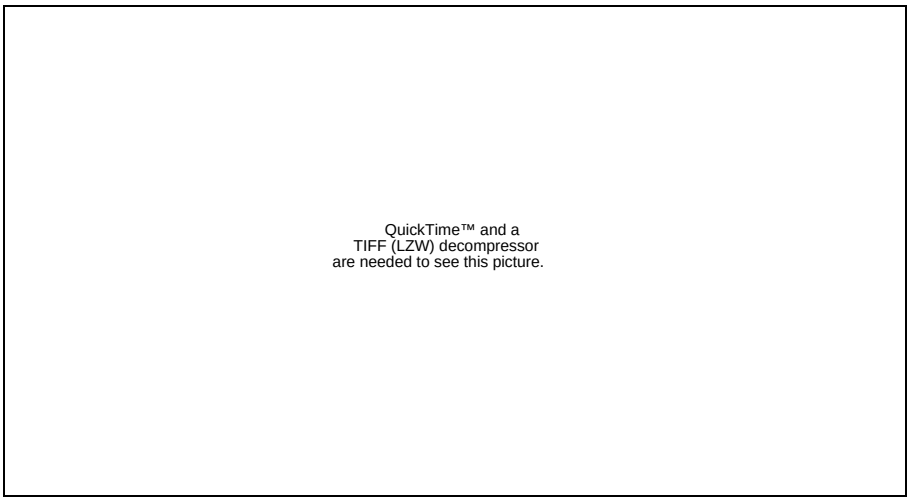

Figure 2

Figure 2 depicts graphically how important the IAR track has grown in relation to the other 40 tracks held at SAC 2006 accounting for a good percentage of the accepted papers.

3 Conclusions

The story of the ACM SAC-IAR track has been a success story, so far. It started almost as a joke, from our coincidental desire to go and visit some friends in a nice city (how many other workshops or tracks have started in that same way?), but it has grown to become a recognised name. In particular, it has become what we had hoped, that is an alternative forum to more traditional ones where IR researchers can present the results of their latest work, even if it is still in progress or it is not theoretically very strong or it has not been evaluated using TREC collections and methodology.

We plan to run the IAR track at SAC for only a few more years and then pass its organisation to other capable hands. We hope we will be able to find some very dedicated souls to take our job and continue this success story.

4 Acknowledgments

We thank all PC members who in these five years have allowed the ACM SAC IAR Special Track to be so successful. They provided a great support in reviewing the papers and their comments have certainly been of great help for authors in improving their works.