ECIR 2012: 34th European Conference on Information Retrieval Research

Ricardo Baeza-Yates
Yahoo! Research, Spain
rbaeza@acm.org

Mari-Carmen Marcos
Universitat Pompeu Fabra
Spain
mcarmen.marcos@upf.edu

Arjen P. de Vries
Centrum Wiskunde &
Informatica, The Netherlands
arjen@acm.org

Hugo Zaragoza
Websays, Spain
hugo@hugo-zaragoza.net

B. Barla Cambazoglu
Yahoo! Research, Spain
barla@yahoo-inc.com

Vanessa Murdock
Microsoft, USA
vanessa.murdock@yahoo.com

Alvaro Barreiro
University of A
Coruña, Spain
barreiro@udc.es

David E. Losada
Universidad de Santiago
de Compostela, Spain
david.losada@usc.es

Ronny Lempel
Yahoo! Labs, Israel
rlempel@yahoo-inc.com

Fabrizio Silvestri
ISTI-CNR, Italy
fabrizio.silvestri@isti.cnr.it

Mounia Lalmas
Yahoo! Research, Spain
mounia@acm.org

1 Overview

The British Computer Society's Information Retrieval Specialist Group's European Conference on Information Retrieval (ECIR) is the main European forum for the presentation of new research results in the field of information retrieval. The conference has been running in various forms since 1979. The most recent editions of the conference were held in Rome, Italy (2007); Glasgow, UK (2008); Toulouse, France (2009), Milton Keynes, UK (2010) and Dublin, Ireland (2011).

The 34th European Conference on Information Retrieval (ECIR 2012) was held at the Pompeu Fabra University from April 1, 2012 to April 5, 2012 in Barcelona, Spain, chaired by Ricardo Baeza-

Yates (Yahoo! Research). The conference was jointly organized by Yahoo! Research Barcelona, Barcelona Media Foundation, and Universitat Pompeu Fabra (UPF), with Mari-Carmen Marcos (UPF) as local arrangements chair. ECIR 2012 invited researchers to submit research papers, posters, and demonstrations. The conference also included an industry track, a tutorial program, and three workshops.

The conference received a total of 261 submissions across four categories: 163 full paper submissions, 78 poster submissions, 11 demonstration submissions and 9 industry track submissions. Naturally, the largest proportion originated from Europe (66%), but Asia (17%) and America (14%) were also well represented. Including the tutorials, workshops and the industry day, the conference attracted more than 200 attendees.

Thank to our sponsors, Google (platinum), Microsoft Research (gold), Yahoo! Labs (gold), and Yandex (silver), we kept the tradition of offering very reasonable registration fees with discounts for students and BCS/SIGIR members, as well as many student scholarships. At the end, in spite of a tight budget, the conference obtained a small surplus.

2 Venue & Social Program

The conference was held in the Ciutadella campus (c/Ramon Trias Fargas, 25-27, 08005 Barcelona) of the Universitat Pompeu Fabra. The campus was located 20 minutes walking distance from the city centre. The main venue of the conference was the Auditorium, a room of 596 m^2, which can accommodate up to 350 delegates. It was equipped with a state-of-the-art infrastructure designed to host this kind of events (with computer projector, speakers, wireless microphones, wi-fi and plugs for the attendants).

Coffee breaks were held in the wide corridor in front of the Auditorium. This corridor had access to the Expositions Room, a space of 735 m^2 where the poster and demonstration sessions were held. Tutorials and workshops were held the day before the main conference at classrooms of the same campus. Coffee breaks and lunch catering were served at the ground.

The conference participants were given tickets for having lunch at the university cafeteria every day. As students were in holidays, the cafeteria was available just for ECIR and few university employees. Lunch consisted of starter, main dish, dessert, bread, and drinks. People could choose among several options every day, and the menu changed on a daily basis.

The welcome reception was held at the conference venue. During the reception, wine, beverages, and light hors d'oeuvres were served. The dinner was held at La Fontana restaurant. The restaurant offered typical Spanish and Catalan tapas and dishes, as well as great Catalan and Spanish wine.

3 Invited Talks

ECIR 2012 included two invited talks:

- Paolo Boldi, Studying network structures for IR: The impact of size.
- Yoelle Maarek, The surprising role of users in Web search.

The first talk, given by Paolo Boldi (University of Milan), presented an overview of some graph-theoretic algorithms that are frequently used in the context of web search and social networks. Boldi pointed that, although these algorithms are well understood, the growth in the size of the graphs emerging in today's practical problems poses new challenges. The talk included a discussion of these challenges, some examples about how to cope with these challenges, and a performance comparison of different graph-theoretic algorithms in large problem settings.

The second talk was on the impact of usage data on Web search, given by Yoelle Maarek (Yahoo! Labs). Maarek provided an overview of the role of the users in commercial web search engines and the key aspects of the interaction between the users and the search engines. The talk also contained a discussion on the limitations and risks of using private user data, together with some potential future research directions.

4 Main Technical Program

The program co-chairs, Arjen P. de Vries (Centrum Wiskunde & Informatica) and Hugo Zaragoza (Websays), decided to not distinguish between short and long papers, but accept any size paper under a maximum page length. The number of 163 paper submissions is slightly down from previous years, which we explain by not including a separate short paper category.

For full paper reviewing, the chairs chose to employ the two-tier reviewing model first explored at SIGIR 2011 (and newly introduced for ECIR): two meta-reviewers supervise together a team of three reviewers, given the assignment to lead discussion until the decision is clear and well supported by the set of reviews. Each paper has an experienced primary meta-reviewer, who can make suggestions to the reviewers when the review has issues (not submitted yet, too short, too vague, contradictory, grades do not correspond to review, claims not supported by references, etc.) and can afterwards steer the discussion between reviewers to try to reach a consensus on the reviewing outcome. In a few cases (review missing, no consensus reached, not enough confidence in reviewers, etc.), the secondary meta-reviewer would be asked to enter also a review. Assigning each paper two meta-reviewers ensured that no papers would fall through the cracks of the system and end up with really bad reviews. For this task, a total of 37 meta-reviewers were hand-picked by the program chairs to try to cover reasonably well most areas of the conference. Then, reviewers were invited to obtain a final set of 217 confirmed reviewers. Since we needed considerably more reviewers than previous years, we tried to extend the pool with "younger" reviewers, by inviting authors that had published at ECIR in the last three years and had at least another publication in a top IR conference.

Eventually, the program committee decided to accept 35 full papers (which corresponds to a 21% acceptance rate). An attractive feature of the finally selected program is that the number of papers accepted allowed us to run a single track, such that most delegates saw most papers being presented, and all presenters got a chance for feedback from the full audience. Judging from discussion with attendees, many people enjoyed the breadth of topics they were exposed in the program. All sessions were fully attended (despite the conference being held in Barcelona) and many live discussions were heard during the ensuing breaks and dinners.

The 35 accepted papers covered all areas of IR and its relations to machine learning, natural language processing, data bases, social media, human computer interaction, etc. These papers

were distributed over 11 sessions, titled Applications, Blog and Online-Community Search, Classification, Categorization and Clustering, Evaluation, Image and Video Retrieval, Query Representation, Retrieval Models (2), Semi-structured Retrieval (2), and Systems Efficiency.

Contributions spanned a wide range of domains and applications, including news, e-mail search, medical information online, tourism, and web people search. Social media and semi-structured information (entities!) were featured in quite a few papers, while evaluation was covered in papers on its foundations (a great talk by Stephen Robertson), metrics for speech retrieval, and crowdsourcing. The program included five excellent papers on retrieval models (including the best paper, briefly introduced below), and two papers exploiting and providing new insights in query logs. Other topics of the conference program included information access to image and video data, classification and clustering, and system efficiency.

An award committee chaired by Nick Belkin assigned the best paper award to "Top-k retrieval using facility location analysis", by Guido Zuccon, Leif Azzopardi, Dell Zhang, and Jun Wang. This paper (with a student first author!) explains how facility location analysis, a method from the operations research field, generalizes state-of-the-art models for diversification, but also suggests a more natural model, where in essence the k best representatives of the relevant documents are selected. This new model then is shown to be a successful method on two TREC diversity collections.

5 Posters & Demonstrations

The poster and demonstration sessions took place on the first day of the main conference. The two tracks were chaired by B. Barla Cambazoglu (Yahoo! Research) and Vanessa Murdock (Microsoft), respectively.

5.1 Posters

The poster track received 78 submissions. Every submission was reviewed by at least three reviewers. After the reviewing phase, 24 submissions were accepted for publication. In addition, upon the recommendation of the program committee, four full paper submissions were also accepted for publication as a poster.

The most notable topics in the presented posters were IR theory and formal models, content representation and processing, Web and social media IR, and evaluation. The poster presentations triggered much discussion and facilitated interaction between the conference attendees.

An award is given for the best poster. To determine the best poster award, an award committee consisting of three researchers was formed. The committee considered six of the accepted posters as candidates for the award. After a thorough review process, the best poster award was given to the poster entitled "Predicting IMDB movie ratings using social media", authored by Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten de Rijke from the University of Amsterdam. The best poster award was announced during the conference banquet, together with the other awards.

5.2 Demonstrations

The demonstration track received 11 submissions, of which seven were accepted. Reviewing was double-blind, and authors were asked to submit a brief two-page description of their demo, which was included in the proceedings. The demos covered a range of technologies, mainly in the context of social media streams and search.

6 Tutorials & Workshops

As in the previous three editions, the main conference was preceded by a workshops and tutorial day. The tutorial and workshop chairs were Alvaro Barreiro (University of A Coruña) and David Losada (University of Santiago de Compostela), respectively.

6.1 Tutorials

Five tutorials were selected by an international committee that agreed on the quality and attractiveness of the proposals. The following provides a summary of the tutorials.

Gianluca Demartini, Peter Mika, Thanh Tran and Arjen P. de Vries, "From expert finding to entity search on the Web" (full day): In this tutorial, the speakers presented an overview of techniques and algorithms for different entity search tasks (expert finding, entity ranking in Wikipedia), and ranking structured data extracted from web pages or published as linked data, showing how core IR and Semantic Web techniques are used to accomplish these tasks. Additionally, results from evaluation studies performed at standard evaluation initiatives were reported.

Saeedeh Momtazi and Dietrich Klakow, "Question answering systems: History and architecture" (half day): This tutorial reviewed the history of question answering systems including the main changes that have been made to improve them from the early stages. The question answering tracks at TREC and CLEF and their evaluation criteria were included in this comprehensive introduction. After that, the tutorial covered the most important components which are required for building a question answering system with a clear and detailed overview of the state-of-the-art techniques and models that are applied on different components.

Benjamin Piwowarski and Massimo Melucci, "Quantum information access and retrieval" (half day): During the tutorial, the speakers detailed what quantum probabilities are and how researchers in information access and retrieval attempt to use them. The tutorial paid a great deal of attention to how the different concepts of quantum probabilities (superposition, subspaces, linear operators, entanglement, interference) that distinguish it from the standard probability theory that has been used in IR and the two "quantum" frameworks that have been developed by the presenters were explained. To emphasize the potential applications, examples of code using the kernel quantum probability API were given.

Tony Russell-Rose, "Designing the search experience" (half day): In this tutorial, Tony explored the fundamental concepts and principles of user-centered design for information search and discovery and demonstrated how to apply them in a range of practical contexts. Participants learnt how to differentiate between various types of search behavior developed an understanding of the key dimensions within the search user experience, and discovered how to apply UI design

patterns to commercial search applications. The session concluded with a group exercise applying these skills to a range of practical design challenges.

Markus Schedl and Peter Knees, "Music information retrieval 2.0" (half day): This tutorial reported on the state-of-the-art in mining music-related information from the Web and further giving the audience an introduction to content-based feature extraction. The presenters gave a sound and comprehensive introduction to exploiting Web and community based media in the music domain. The presented approaches are highly valuable for tasks and applications such as automated music playlist generation, personalized Web radio, music recommender systems, and intelligent user interfaces to music.

6.2 Workshops

ECIR 2012 included an open call for workshops to stimulate researchers to propose exciting and dynamic workshops that have significant scope for participant interaction. The received workshop proposals were reviewed by three referees from the workshop program committee and, finally, three proposals were accepted. Two workshops were scheduled as half-day workshops and another one was a full-day workshop. The day of the workshops was vibrant and the planned activities attracted the attention of many ECIR participants.

The "Searching 4 Fun!" Workshop (full day) was organized by David Elsweiler, Morgan Harvey and Max L. Wilson. The workshop was focused on pleasure-driven search and contained discussions on different challenges in the increasingly important area of casual-leisure search. The intention, motivation, and querying behavior behind this type of search activities differ from typical web search and the workshop aimed to bring together relevant communities (e.g. recommender systems, result diversity, multimedia retrieval) to discuss new and early research and create a vision for future work in this area. The workshop included a keynote speech by Elaine Toms ("Finding without seeking, retrieving without searching") and nine exciting papers. The workshop was run by the organizers in a highly interactive manner and the sessions were very lively.

Birger Larsen, Christina Lioma and Arjen P. de Vries organized the "Task-Based and aggregated search" workshop (half day). The workshop aimed to stimulate exploratory research in task-based and aggregated search, and to investigate synergies between these two areas. Research into task-based search aims to understand the user's current task and desired outcomes, and how this may provide useful context for the information retrieval process. The workshop's program included a keynote by Kalervo Jarvelin ("Information access and integration in task-based heterogeneous environments"), two paper sessions where a total of six papers were presented, a poster session with two posters, and a final interactive discussion. A best paper award was given to Thomas Beckers and Norbert Fuhr for their contribution entitled "Towards the systematic design of IR systems supporting complex search tasks", and a best poster award was given to Frank Sawitzki, Philipp Schaer and Daniel Hienert for their poster entitled "Extending aggregated search in a social sciences digital library".

The "Information retrieval over query sessions" workshop (half day), which was jointly organized by Ben Carterette, Paul D. Clough, Evangelos Kanoulas and Mark Sanderson, focused on search scenarios where users issue multiple queries that form part of a search session involving a sequence of user-system interactions. This might be the result of users clarifying or refining their information needs (e.g., broadening or narrowing a search), in response to the search results (e.g.,

too few or too many hits). SIR2012 aimed to provide discussion and promote research and development on three main themes: retrieval models and ranking, evaluation and test collections, and user interaction and interfaces. The workshop included talks by Mounia Lalmas ("Session, focus, and engagement"), and Gene Golovchinsky ("Exploring session search"), and the presentation of five papers.

7 Industry Day

The ECIR 2012 Industry Day was held on April 5, 2012. Its program was chaired by Ronny Lempel (Yahoo! Labs) and Fabrizio Silvestri (ISTI-CNR). The speakers, mainly coming from companies in the area of search and information retrieval, presented industrial research findings. The final program included three invited talks and six peer-reviewed contributions, and was attended by about 60 people.

The first invited speaker, Arjen de Vries, showed the research behind Spinque, commercialized by a technology spin-off from Centrum door Wiskunde & Informatica, CWI (the Dutch National Research Centre in Computer Science and Mathematics). Spinque's technology is based on a search-by-strategy paradigm that consists of "instantiating" different search engines depending on the clients requirements. The second invited speaker, Alexandros Karatzoglou from Telefonica R+D, presented a recommender system designed by Telefonica that is particularly suited for mobile devices and leverages the context of users. The final invited speaker of the day was Paolo Ferragina (University of Pisa), who explained how his research group is turning research results into an industrial product by means of vertical applications built on top of TAGME, a framework for clustering, classification and similarity-comparisons of short texts.

The main theme of the contributed talks was text analytics, from innovations in the platform (Websays and AOL) through applications of synonym mining for eCommerce search (eBay), to the power of sentiment analysis (Gavagai, AOL). In a different vein, novel interfaces for exploring social media were presented by Oxyme. An interesting point highlighted by all speakers was the difficulty of managing and processing streams of dynamic web data. In particular, the speakers agreed that large amounts of editorial work are needed to effectively enable and tune their proposed techniques.

8 Student Mentoring Program

ECIR 2012 offered a mentoring program, chaired by Mounia Lalmas (Yahoo! Research). The program aimed to help students networking with senior researchers during the conference by linking them up with appropriate people beforehand. Each student attending the conference was assigned a senior researcher to discuss his or her PhD work/research/paper. Students with a paper, poster, or demo were automatically assigned a mentor. Others were assigned a mentor upon request, by providing a short abstract of their PhD work.

The task of the mentors was to provide feedback on the student paper/PhD topic during one of the various coffee breaks and social events, or on any specific research issues students wished to discuss. Mentors and students made contact at some point before or during the conference, to arrange for a meeting time to discuss the student research interests.

34 senior researchers from academia and industry, throughout the conference, mentored a total of 34 students. Many of the mentors volunteered themselves through a call for participation, and great thanks go to them. The feedback was positive, from both mentors and mentees, and it was a pleasure to see mentors and mentees talking, and learning, and sometimes from both sides. It was a particular delight to see so many students working in information retrieval and eager to learn more, and to become part of a large network of researchers working in information retrieval and related areas.

Acknowledgments

We would like to thank the publicity chair, Roi Blanco, the registration chair, Aristides Gionis, the proceedings chair, B. Barla Cambazoglu, and the overall organizing committee. Finally, we specially thank again our sponsors, Google (platinum), Microsoft Research (gold), Yahoo! Labs (gold), and Yandex (silver).