ECIR 2013: 35th European Conference on Information Retrieval

Pavel Serdyukov
Yandex
pavser@yandex-team.ru

Pavel Braslavski
Ural Federal University
& Kontur Labs
pbras@yandex.ru

Jaap Kamps
University of Amsterdam
kamps@uva.nl

1 Overview

The European Conference on Information Retrieval (ECIR) is the main European forum for the presentation of new research results in the field of information retrieval and cognate areas. Initiated as a BCS Information Retrieval Specialist Group (IRSG) Colloquium for Information Retrieval in 1979, the event developed in a renowned international scientific meeting. To reflect its European scale, since 1998 the conference was held alternately in the UK and continental Europe. From 2012, this rule has been removed. The most recent ECIR editions were held in Glasgow, UK (2008), Toulouse, France (2009), Milton Keynes, UK (2010), Dublin, Ireland (2011), and Barcelona, Spain (2012).

The 35th European Conference on Information Retrieval took place March 24-27, 2013 in Moscow (Russia). The conference was jointly organized by Yandex and Higher School of Economics (HSE). The conference became the easternmost ECIR ever.

ECIR 2013 in Moscow marks an important step in the development of the conference series: geographic expansion, a high number of participants, and the largest number of accepted papers (which resulted in an impressive 900-page proceedings volume). In addition, the Moscow edition of the conference introduced some new features: a two-tier program committee, a redesigned mentoring program, a boaster session for poster presentations, and attracted many new participants.

ECIR 2013 received a total of 287 submissions in three categories: 191 full papers, 78 posters, and 18 demonstrations. The geographical distribution of the submissions is as follows: 70% were from Europe (including 9% from Russia), 17% from Asia, 12% from North and South America, and 3% from the rest of the world. All submissions were reviewed by at least three members of an international two-tier Program Committee. Of the papers submitted to the main research track, 30 were selected for oral presentation and 25 for poster/short presentation (16% and 13% respectively, hence a 29% acceptance rate). In addition, 38 posters (49%) and 10 demonstrations (56%) were accepted. Out of accepted contributions 66% have a student as the primary author. Additionally, ECIR 2013 hosted four tutorials and two workshops covering various IR-related topics.

2 Venue & Social Program

Moscow welcomed conference participants with freeze and snow storm that were exceptional for this time of the year – this March was recorded in Moscow as the coldest one in the last 33 years. However, these circumstances did not discourage conference attendees; many of them were quite excited to experience Russian snowy winter in late March.

Tutorials and workshops were held on Saturday, 24th March at the Yandex headquarters. Welcoming reception took place in the evening in the same place. Wine and hors d'oeuvres, including traditional Russian dishes and drinks, were served for participants. During the welcoming reception the participants had a great opportunity to establish new contacts and colleagues already acquainted with each others could catch up on the latest news.

The main conference (25–27 March, 2013) was held at the Digital October Center[1], located in the very heart of Moscow in the buildings of the former Red October Chocolate Factory, on Bolotny Island in the middle of Moskva River. The conference center is located in a trendy area of Moscow, where media and technology companies, advertising agencies, as well as restaurants and night clubs are also located. Digital October Center has been a host for many high-level events devoted to information technology. Conference venue was located within walking distance from the main attractions in the city center – Red Square, the Cathedral of Christ the Saviour, Pushkin museum of fine arts, and others. The main conference hall of 350 m2 accommodated around 350 attendees and was equipped with a large TV wall (see Figure 1), additional wide TV screens, wireless microphones, wifi, etc.

Figure 1: Main conference presentation screen with the conference logo

Poster and demonstration sessions were held on Monday evenining in the wide corridor and the adjacent rooms of the venue. Wine and snacks were served to heat the discussions, which ended far after the planned time.

Every day, coffee breaks and snacks were served in the area around the main conference hall. Participants were also provided with free lunches (a variety of sandwiches and burgers) at each day of the main program.

[1] `http://digitaloctober.ru/en/`

The conference dinner was held on the evening of Tuesday, 26th March, on a yacht called Butterfly (see Figure 2) that cruised through the Moskva River. Food and wine were served aboard. Conference participants could enjoy the picturesque views of Moscow at night through huge panoramic windows. Due to weather conditions the yacht could even demonstrate its ice-breaking capacity. Following the tradition of most information retrieval conferences, the awards for the best paper, best student paper and best short paper were also announced and celebrated at the dinner (see Figure 4)

Figure 2: Floating banquet venue

3 Keynotes

ECIR 2013 included two keynote talks as well as seven invited talks in the Industry Day (see Section 6 for details). Mor Naaman gave an inspiring opening keynote on social media information about events. Diane Kelly gave the "Karen Spärck-Jones Award" keynote on the past, present and future interactive IR research practices, and how our research is often constrained by the systems we are familiar with at a given point in time.

3.1 Time for Events: Telling the World's Stories from Social Media

Mor Naaman (Rutgers University, currently at Cornell Tech NYC, and Mahaya Inc) gave the opening keynote, entitled "It's Time for Events: Telling the World's Stories from Social Media."

An overwhelming amount of information from real-world events is shared by individuals through social media services like Facebook, Twitter, Instagram and YouTube. These events range from major global events like an uprising or an earthquake, to local events and emergencies such as a fire or a parade; from international media events like the Oscars, to events that enjoy little media coverage such as a conference or a music concert. This shared media represents an important part of our society, culture and history. At the same time, this social media event content is currently fragmented across services, hard to find, and often difficult to consume due to its sheer scale. Mor

Figure 3: ECIR keynote speakers: Mor Naaman and Diane Kelly

tackled three critical challenges in making social media information about events accessible and usable: 1) the detection of events in social media content, 2) identification and ranking of content relevant to an event across social media sites, and 3) organization and presentation of event data to allow users to effectively explore, analyse, and experience an event through its social media content.

Mor described research on the organization and presentation of social media event content. He described two systems built to help journalists make use of social media data around events; one helps journalists asses the public response to large-scale media events (e.g. a presidential speech); another assists them in finding knowledgeable on-the-ground sources for breaking news events. Second, he discussed a system designed to improve the consumption and experience of viewing videos recorded and shared at real world events. This results in new tools that allow multiple stakeholders, such as journalists, first responders, researchers, policy makers and the general public, to see and understand the stories of world, as told in social media.

Mor was introduced as the perfect ECIR keynote fitting the impossible combination of appealing to computer science and information science, to theory and system building, to academia and industry, to the past of classic IR and the future of modern Web, etc. He certainly lived up to these impossible expectations: the broad and positive, forward looking focus of the keynote resonated through the discussions during the rest of the conference. A video registration of the keynote is online at `http://www.youtube.com/watch?v=cikg5X304Xo`.

3.2 Contours and Convergence

Diane Kelly (University of North Carolina, Chapel Hill) gave the Karen Spärck-Jones Award Talk, entitled "Contours and Convergence: Past, Present and Future of (Interactive) IR Research Practices."

Diane gave a wonderful overview of the history of "User" studies over the years, emphasizing interactive IR and therefore user interfaces that support interactive search. This started with the early work from the 1960s and 1970s, such as the early work on relevance feedback in the 1960s, and the wonderful work on advanced interfaces for bibliographic search as pioneered by Donald Walker and others in the early 1970s. The 1980s saw important work on user modelling and personalization, as well as first graphical user interfaces. The adoption of graphical interfaces

gave endless new possibilities for user interface design, and in the 1990s this led to many highly original designs – such as tile bars, visualization tools, clustering tools, hierarchical arrangements, newspaper or spatial memory metaphors. By far the 1990s was the most productive decade in terms of innovative user interface design (although not every wild idea survived the test of time). The 2000s saw a demise of explicit feedback and anything else involving user effort, and a strong focus on implicit feedback, large-scale log analysis and machine learning, all happening under the hood of the system, hidden from the users. The overview wonderfully illustrated how inspiring the 1990s was for researchers to think outside the box and experiment with completely new concepts, and how difficult it is nowadays to do any research that isn't framed by the current conventions of web search dominating a large part of research and of our daily lives.

Diane outlined some contours of the present and future. The first two addressed experimental research with users in the lab: Expect more from people (i.e., stop treating them like lazy idiots!). Stop asking people if they are satisfied (i.e., we need new tools and methods to collect interaction data, and to understand user variance). Two of them build new bridges to the system-centered research in IR: Develop formal languages for modeling search behavior and information use. Explore new statistical techniques for modeling behavior, interactions and user experience. The final two – slow down; read more, experiment less – addressed the change in research practice over the years, with a exhausting pace nowadays that may lead to low impact or even flawed research.

Diane's talk contained a range of quotes from Karen's papers (in particular from the 1988 SIGIR Salton Award paper), highlighting that some of the main challenges remain still unsolved up to this day. We warmly thank the Karen Spärck-Jones Award Panel for contributing such an outstanding keynote talk to the ECIR 2013 program. Her talk touched the audience in a special way by making them realize how framed our own work and our own thinking is by the types of systems we usually deal with. A video registration of the keynote is online at `http://www.youtube.com/watch?v=sv7_d23UrXQ`.

4 Main Technical Program

ECIR received a total of 287 submissions in three categories: 191 full paper submissions, 78 short paper submissions for poster presentation, and 18 short paper submissions for demo presentation. Note that we clearly distinguished between the peer reviewed publication of a short paper (up to 4 pages) in the conference proceeding from the mode of presentation during the conference (either poster or demo presentation).

4.1 Program Committee

As one of the program committee chairs remarked at the opening talk: ECIR is too large to be a small conference. Hence the technical program of ECIR requires reviewing procedures and quality assurances that are on par with the other top conferences in the field.

The program committee chairs continued the successful two-tier model of ECIR 2012, with two senior program committee members to supervise together a team of three reviewers, given the assignment to lead the discussion until the decision is clear and well supported by the set of reviews. Each submission had an experienced primary SPC member assigned, who supervised the reviews (signalling missing, incomplete or unsupported reviews), led the discussion (trying to reach

consensus or a group decision on the reviewing outcome and its motivation), and wrote a final meta review summarizing the outcome and making recommendations. The secondary SPC member was following the discussion, and stepped in when the need required it. In some cases the secondary SPC acted as additional reviewer to step in for a missing review, or to give a fresh eye when it was difficult to reach consensus amongst the reviewers. The conference management software was adapted to support the dual SPC assignments, all discussion happened in the conference system, often leading to extensive threads of arguments.

A large PC was recruited with the aim to have a reasonable workload for all PC and SPC members. There were a total of 41 SPC members, who each managed on average 4.7 papers in the role of primary SPC member, and again 4.7 papers in the role of secondary SPC member. This workload allowed SPC members to devote sufficient time to each paper under their care, for both the primary and secondary assignments. The PC consisted of 190 PC members, with on average 3 full papers to review and in addition 1.5 short paper. Again, a relatively low workload that certainly contributed to the quality and depth of the reviews and subsequent discussions.

The program committee decided before reviewing started that we would not "demote" full papers to posters – which would have required authors to significantly reduce their papers (with likely some of the best parts falling out). Instead, we planned to accept those papers also as full papers in the proceedings, but have an alternative short oral presentation at the conference. This affected some of the reviewing parameters, making reviewers more critical in general. Reviewers tended to find the second category very attractive for strong papers that had some issues (e.g., a very strong paper that primarily extends earlier published results hence was consider less novel) and became more critical of recommending accept with long oral presentation. Also papers that had interesting ideas, but issues with clarity, language, writing, or structure, were less easy given an accept with short presentation – arguing against the risk of accepting sub-standard papers into the proceedings.

The PC chairs oversaw the discussion and progress, and critically discussed the draft recommendations of the primary SPC member in the meta review (both in terms of scores and in terms of the review text, and considering the meta-review, first tier reviews, as well as the discussion). This led to a draft program, that was again critically discussed on potential biased in favour of, or against, particular types of research or student versus senior authors. All this discussion led to relatively few changes, and most changes concerned borderline cases, and overall the two-tier system functioned remarkably very well. A crucial factor is the reviewer assignment, which was done with considerable care and aided by a large PC and SPC with many reviewers for any area of expertise. Both PC and SPC members did bid on submitted abstracts, and the assignment was hand made by one of the PC chairs based on their bids and known expertise and prior reputation as a reviewer, while balancing the whole team assigned to each paper in terms of senior and junior reviewers.

Although ECIR was once known as a student conference, but the reviewing procedure did not privilege student authors in any special way over more senior authors. As it turned out a large fraction of the submitted and accepted papers had a student main author, both for papers from academia and for papers from industry. See Table 1 for details of the submitted and accepted papers, which we will discuss in the sections directly below.

The regional distribution (based on the contact authors of all 103 accepted submissions) is as follows: 70 papers (or 68%) is from Europe, including 9 papers (9%) from Russia; 18 papers

Table 1: ECIR 2013 acceptance rates

	Submitted	Accepted		Student	
Full paper (oral)	191	30	16%	21	70%
Full paper (short)		25	13%	16	64%
Full paper (combined)		55	29%	37	67%
Short paper (poster)	78	38	49%	28	74%
Short paper (demo)	18	10	56%	3	30%
All	287	103	36%	68	66%

(17%) are from Asia; 12 papers (12%) is from the Americas; and the remaining 3 papers (3%) from the Middle East. Hence, there is a highly international audience with a (non-surprising) clear European focus at the European Conference on IR.

The above resulted in a strong technical program with a record number of papers (103) and a proceedings with a record number of pages (920) [5].

4.2 Full papers

The final outcome was that the program committee accepted 30 of the 191 full paper submissions as full paper with oral presentation (a 16% acceptance rate). While this is a much lower percentage as last year (21%) this is mainly caused by the increased number of submissions (ECIR 2012 accepted 35 full papers). Other 25 papers (13% of the submissions) were accepted in full in the proceedings, but with a short oral presentation at the conference. Hence, overall, a total of 55 full papers appeared in the proceedings (an overall acceptance rate of 29%).

The 30 papers in the first category (30 minutes oral presentation) were presented in 10 sessions of three papers each: User Aspects, Multimedia and Cross-media IR, Data Mining, IR Theory and Formal Models, IR System Architectures, Classification, Web, Event Detection, Temporal IR, and Microblog Search. The 25 papers in the second category were all presented in a single plenary session (5 minutes each) – clearly one of the most exciting sessions of the whole conference. The short oral presentations were followed by a poster presentation during the Posters and Demos session.

4.3 Short Papers with Poster Presentation

The short papers with poster presentation were chaired by Eugene Agichtein. Short papers were up to four pages in Springer's LNCS style. We received a total of 78 submissions, and each submission was reviewed by three reviewers of the same program committee that reviewed the full papers. There was no SPC assigned, but the poster chair oversaw the reviewing and discussion, aided by one of the PC chairs. As it turned out, a total of 38 short papers were accepted for poster presentation (a 49% acceptance rate).

4.4 Short Papers with Demo Presentation

The short papers with demo presentation were chaired by Emine Yilmaz. Short papers were up to four pages in Springer's LNCS style. We received a total of 18 submissions, and each submission was reviewed by three reviewers of the same program committee that reviewed the full papers. There was no SPC assigned, but the poster chair oversaw the reviewing and discussion, aided by one of the PC chairs. As it turned out, a total of 10 short papers was accepted for demo presentation (a 56% acceptance rate).

Demo submissions were *not* anonymized to allow the authors to include a link to the actual system to demonstrate at the conference – following a suggestion of the ECIR'12 demo chair, Vanessa Murdock. This greatly facilitated the reviewing of demonstrations.

4.5 Best Paper Award

Figure 4: Best paper, best student paper and best short paper awards winners

A best paper awards committee chaired by Arjen de Vries, and further consisting of Gabriella Pasi, Leif Azzopardi, Fabrizio Silvestri, and Birger Larsen, selected winners of the Best Paper Award (for full papers), the Best Student Paper Award (again for full papers), and the Best Poster Award (for short papers). Nominations were provided by one of the PC chairs, for both the full paper and short paper submissions, a set of papers was selected by merging the sets of the three highest ranked papers, the three highest ranked papers that had a student author, and the three highest ranked papers that received a "best paper award" recommendation.

The ECIR 2013 Best Paper Award was presented *ex aequo* to two paper that received equal support within the best paper awards committee: *Semantic Search Log k-Anonymization with Generalized k-Cores of Query Concept Graph*, by Claudio Carpineto and Giovanni Romano [1]; and *Understanding Relevance: An fMRI Study*, by Yashar Moshfeghi, Luisa Pinto, Frank Pollick, and Joemon M. Jose [4].

The ECIR 2013 Best Student Paper Award was presented to: *Using Intent Information to Model User Behavior in Diversified Search*, by Aleksandr Chuklin, Pavel Serdyukov, and Maarten de Rijke [2].

The ECIR 2013 Best Poster Paper Award was presented to: *Optimizing nDCG Gains by Minimizing Effect of Label Inconsistency*, by Pavel Metrikov, Virgil Pavlu, and Javed Aslam [3].

5 Tutorials and Workshops

5.1 Tutorials

Djoerd Hiemstra chaired the tutorial program, and four half-day tutorials were selected from ten proposals submitted.

5.1.1 Searching the Web of Data

Gerard de Melo and Katja Hose gave a tutorial on "Searching the Web of Data." Search is currently undergoing a major paradigm shift away from the traditional document-centric "10 blue links" towards more explicit and actionable information. Users expect the system to "understand" the user's information need and respond to it more directly instead of only serving documents matching the given set of keywords. Recent advances in this area are Google Knowledge Graph, Virtual Personal Assistants such as Siri and Google Now, as well as the now ubiquitous entity-oriented vertical search results for places, products, etc. Apart from novel query understanding methods, these developments are largely driven by structured data that is blended into the Web Search experience. Structured data can be obtained from a wide variety of documents and Web sources by tapping on information extraction and semantic markup like microformats. Additionally, the Web already offers publicly accessible knowledge bases, such as DBpedia, Yago, Freebase and the Linked Open Data cloud, as also used in Google's Knowledge Graph. Providing vast amounts of information about many different types of entities, these data sets can become very large, so sophisticated techniques for organizing and querying them are required. We discuss efficient indexing and query processing techniques to tackle these challenges. Finally, we present query interpretation and understanding methods to map user queries to these structured data sources, also highlighting the recent trend of virtual personal assistants like Siri.

5.1.2 Distributed Information Retrieval and Applications

Fabio Crestani and Ilya Markov gave a tutorial on "Distributed Information Retrieval and Applications." Distributed Information Retrieval (DIR) is a generic area of research that brings together techniques, such as resource selection and results aggregation, dealing with data that, for organizational or technical reasons, cannot be managed centrally. Existing and potential applications of DIR methods vary from blog retrieval to aggregated search and from multimedia and multilingual retrieval to distributed Web search. In the first part of the tutorial we will briefly discuss the main DIR phases, that are resource description, resource selection, results merging and results presentation. In particular, the attendees will get familiar with the ways of building high level descriptions of searchable collections. The large and the small document approaches to resource selection will be presented as well as the classification-based approach. The main score normalization and results merging techniques will also be discussed. The first part of the tutorial will be concluded by discussing the ways of presenting search results, coming from multiple sources, to a user. The second and the main part of the tutorial will be dedicated to various applications of DIR methods. In particular, we will discuss blog, expert and desktop search as special instances of the resource selection problem. We will then talk about the rapidly developing area of aggregated search, discussing such problems as vertical selection and results aggregation. Other applications,

such as multilingual and multimedia retrieval, personal meta-search and aggregated Web search, will also be mentioned. We will conclude our tutorial by presenting potential applications of DIR techniques, such as distributed Web search, enterprise search and aggregated mobile search.

5.1.3 Practical Online Retrieval Evaluation

Filip Radlinski and Katja Hofmann gave a tutorial on "Practical Online Retrieval Evaluation." Online evaluation is an evaluation technique that allows techniques developed in the information retrieval community to be assessed based on how real users actually respond to improvements made. Because this technique is directly based on observed user behavior, it is a promising alternative to traditional offline evaluation, which is based on manual relevance assessments, especially in settings where reliable assessments are difficult to obtain (e.g., personalized search) or expensive (e.g., search by trained experts in specialized collections). Despite its advantages, and its successful use in commercial settings, online evaluation is rarely employed outside of large commercial search engines due to a perception that it is impractical at small scales. The goal of this tutorial is to show how online evaluations can be conducted in such settings, demonstrate software to facilitate its use, and promote further research in the area. We will also contrast online evaluation with standard offline evaluation, and provide an overview of online approaches.

5.1.4 Cross-Lingual Probabilistic Topic Modeling

Marie-Francine Moens and Ivan Vulic gave a tutorial on "Cross-Lingual Probabilistic Topic Modeling and its Applications in Information Retrieval." Cross-lingual topic models are a fairly novel group of unsupervised, language-independent and generative machine learning models that can be effectively trained on a large-volume of non-parallel, comparable multilingual data (e.g., multilingual Wikipedia or news data discussing the same events). They offer an elegant way to represent content across different languages. Their probabilistic framework allows for their easy integration into a language modeling framework for cross-lingual information retrieval. The half-day tutorial will give an overview of recent advances in cross-lingual topic modeling and retrieval. It includes: (1) A high-level overview of the key intuitions and assumptions behind topic modeling in general and cross-lingual topic modeling in specific; (2) The methodology and mathematical foundations; and (3) The application of these models in various cross-lingual tasks, with a special focus on cross-lingual information retrieval models. The tutorial first introduces the concept of probabilistic topic modeling, starting from monolingual contexts, where we introduce the key intuitions and describe the most prominent monolingual models such as probabilistic semantic analysis (pLSA) and latent Dirichlet allocation (LDA). We then present a representative cross-lingual topic model called bilingual LDA (BiLDA). We explain its generative story, its training techniques (variational inference and Gibbs sampling) and its inference procedure on unseen text documents. Finally, an important part of the tutorial focuses on the applications of the cross-lingual topic models, where the emphasis is on cross-lingual retrieval models. We also present how to use the knowledge from the models for the tasks of cross-lingual event clustering, cross-lingual document classification and cross-lingual semantic similarity of words.

5.2 Workshops

Evgeniy Gabrilovich chaired the workshop program, and two workshops were accepted from four submissions. In addition, another workshop was conceived after the ECIR deadlines, and was organized in co-location with the conference.

5.2.1 Integrating IR technologies for Professional Search

Michail Salampasis, Norbert Fuhr, Allan Hanbury, Mihai Lupu, Birger Larsen and Henrik Strindberg organized a workshop on "Integrating IR technologies for Professional Search." Many facets of IR technology (e.g. exploratory search, aggregated search, federated search, task-based search, IR over query sessions, cognitive IR approaches, HCIR) aim to at least partially address these demands. To facilitate integration of the results of all these IR technologies a generalised framework is needed which will allow to address not only how to aggregate or merge content from multiple sources or specialised search services, but will also focus on the needs of professional workers using search systems, and how they need to interact with multiple search tools and UIs and use different IR technologies to cope with the complexity of retrieving, finding, understanding and analysing information in their workplace. This workshop aimed to stimulate exploratory research and to bring together various facets of IR research and to promote the discussion between researchers towards the development of a generalised framework facilitating the integration of IR technologies and search tools into next generation professional search systems. This envisioned framework should be supported from new or the extension of existing protocols (e.g. OpenSearch protocol) and may influence the design of next generation professional search systems.

5.2.2 Group Membership and Search

Ingmar Weber, Djoerd Hiemstra and Pavel Serdyukov organized a workshop on "From Republicans to Teenagers: Group Membership and Search (GRUMPS)." In the early years of information retrieval (IR), the focus of research was on systems aspects such as crawling, indexing, and relevancy ranking. Over the years, more and more user-related information such as click information or search history has entered the equation, creating more and more personalized search experiences, though still within the scope of the same overall system. Though fully personalized search is probably desirable, this individualistic perspective does not exploit the fact that a lot of a user's behavior can be explained through their group membership. Children, despite individual differences, share many challenges and needs; as do men, Republicans, Chinese or any user group. This workshop took a group-centric approach to IR and invited contributions that either (i) propose and evaluate IR systems for a particular user group or that (ii) describe how the search behavior of specific groups differ, potentially requiring a different way of addressing their needs. For practical reasons, the workshop was merged with the workshop on "Integrating IR technologies for Professional Search."

5.3 Formal Concept Analysis meets Information Retrieval

Claudio Carpineto, Sergei O. Kuznetsov, and Amedeo Napoli organized a co-located workshop on "Formal Concept Analysis (FCA) meets Information Retrieval." Formal Concept Analysis (FCA)

is a mathematically well-founded theory aimed at data analysis and classification, introduced and detailed in the book of Bernhard Ganter and Rudolf Wille, *Formal Concept Analysis*, Springer 1999. The area came into being in the early 1980s and has since then spawned over 10,000 scientific publications and a variety of practically deployed tools. FCA allows one to build from binary data – a binary table with objects in rows and attributes in columns – a taxonomic data structure called concept lattice which can be used for many purposes, especially for Knowledge Discovery and Information Retrieval. The workshop was intended, on the one hand, to attract researchers from FCA community to a broad discussion of FCA-based research on information retrieval, and, on the other hand, to promote ideas, models, and methods of FCA in IR community. It included papers on information retrieval understood in a rather general sense, including such issues as ontologies, semantic web, text mining, digital libraries, NLP, and related aspects. The workshop addressed two main issues: First, how can FCA support IR activities including but not limited to query analysis, document representation, text classification and clustering, social network mining, access to semantic web data, ontology engineering. Second, how can FCA be extended to address a wider range of IR activities, possibly including new retrieval tasks.

6 Industry day

The ECIR 2013 Industry Day was held on Wednesday, March 27, 2013, during the regular conference program, and in parallel with the technical tracks. The Industry Day's goal was to present the state of the art in search and search-related areas, delivered as keynote talks by influential technical leaders from the Industry. Its program was chaired by Ilya Segalovich, CTO of Yandex, and included seven keynotes.

The first speaker, Paul Ogilvie, Software Engineer at LinkedIn, in his talk "Lessons from the Wild: How Context Can Shape Consumption in Content Recommendation Systems" gave an overview of different kinds of problems his team faces when they try to build recommender systems and evaluate them. Particularly, he mentioned the tasks of recommending jobs and professional groups using features mined from the user profiles, but the talk was mainly focused on the news and blog posts recommendation task. So, the first part of Paul's talk was about different ways they evaluate the performance of that task, including measuring the relevance and the quality of recommendations online, what gives a good sense of whether people are engaging with the system, and doing Cranfield-like and log-based offline evaluations. Special attention has been paid to the biases that need to be considered for online evaluations: presentation, position and content novelty biases, and the variation of activity level over users and the time of the day. The second part of the talk was about taking those biases into account for learning offline to recommend more relevant content using the contextual-bandit-based approach.

Jimmy Lin from Twitter/University of Maryland presented the second keynote "Search and Discovery at Twitter", which overviewed the real-time infrastructure of Twitter intended to deal efficiently with fast, high velocity data streaming into their platform constantly (400+ million tweets per day). Jimmy started from presenting the challenges for such an infrastructure: the need to handle low-latency updates of the tweets collection and high-throughput real-time processing of user queries meaning matching them to tweets by taking into account the fine-grained temporal model of their relevance. Then the talk continued with presenting the general architecture of Twitter search engine, with the core retrieval engine, called EarlyBird, at its center. EarlyBird

is built on Lucene with real-time extensions, covered by the rest of the keynote. Particularly, Jimmy explained different ways of posting lists organization that help to efficiently retrieve relevant tweets in chronological order, described EarlyBird's memory allocation policies and its concurrency management strategy.

Marc Najork form Microsoft Research opened the next talk in the program with the introduction into the social search as it is viewed by Bing search engine. First, he presented a few typical scenarios of using a search engine with a social twist: people search, including the search for celebrities, search for travel recommendations, and local businesses. A web search engine becomes social-aware, when it is informed about different social identities of the user and has access to her profiles at different social platforms, including her social connections. A typical way to improve the standard "impersonal" ranking of documents in such cases is to promote content that has been liked, retweeted or favored in any other way by the user's friends. Since, friendship is not necessarily a binary relation between any two users and can be transitive and characterized by its strength, the second part of the keynote focused exactly on measuring that degree by computing the distance between users in a social network efficiently and with minimum probing of social platforms for additional information. The keynote was concluded with the attempt to point out the limitations of social search and dispel its myths, such as the belief that every search engine user has a social network account, or that users situated close in a social network have similar tastes and personalities.

The fourth talk of the day was given by Antonio Gulli from Bing, UK, who drew the audience's attention to the stack of technologies behind Bing's query suggestion services, such as detecting the dominant query intent, learning to rank suggestions using a range of features while evaluating several target functions and optimizing not only for relevance, but also for freshness and query popularity. Antonio also presented his view on the importance of long and short-term personalization of query suggestion. In the second part of the talk, Antonio described Cosmos, the platform for large-scale data management and mining, that allows to process and analyze huge amounts of data by means of an SQL-like language (Scope) and a more flexible variant of the Map-Reduce parallel computing model.

The program continued with Karen Church from Telefonica Research and her keynote "Mobile search: a force to be reckoned with!". She started with a nostalgic look back at the simple beginnings of mobile search and discussed how, why and in what ways mobile search has evolved over the past 8 years. She highlighted patterns of mobile search usage and demonstrated how they are continually evolving and different from desktop search. Particularly, she stated the fact that mobile phones in contrast to desktops are considered by users as highly personal devices, almost never shared with anyone. Besides, mobile phones can be accessed almost immediately in 90% of life situations, so, basically, almost anytime and anywhere. These observations indicate new opportunities for search algorithms and their stronger impact on users (e.g. 70% of mobile searches lead to action within 1 hour), but also represent certain challenges. Since, it is predicted that by the end of 2013, one of every three queries will come from a mobile device, these challenges cannot wait to be faced and tackled. Later, Karen presented some user-centric studies that evidence the unique needs, intents and motivations of mobile searchers. They also disapprove some of the popular misconceptions, such as, that mobile queries are much shorter (not really shorter and their length is growing), that mobile search is less successful (the level of query abandonment is already comparable to that of desktop search), that queries are largely with local intent (only

1.7% more local queries are observed on smartphones than on desktops), and others. Finally, Karen shared some thoughts about where mobile search is heading and about the challenges that lie ahead on its way toward being able to provide a truly special search experience without being just a watered down version of web search.

The last two talks of the program were given by representatives of two largest Internet companies in Russia: Yandex and Mail.ru. Alexey Voropaev from Mail.ru demonstrated the need for active learning to rank documents and presented a simple and effective learning algorithm that allows to significantly reduce the training set size as well as improve the ranking quality. The primary challenge of active learning is optimal selection of samples for labeling with the aim to maximize the quality of a ranker with every next labeled sample. Alexey presented several possible approaches to iterative selection of those samples, such as using the votes of several algorithms (committees) trained differently to select the right samples, clustering samples with self-organizing maps to identify dense and sparse regions in the training space and limit sampling from very dense regions and, at each step, select samples that cover the maximum number of uncovered clusters to increase diversity of the training set.

Andrey Styskin, the head of web ranking team at Yandex, concluded the Industry day with the keynote "Aggregate and conquer: Finding the way in the diverse world of user intents". He described the approach adopted by Yandex to diversify search results and aggregate them from different verticals into one search engine result page (SERP). Instead of fixing slots on SERP for verticals and their documents, the approach described in most papers on the topic, Yandex adopts a more flexible technique of blending vertical results together, that is based on a probabilistic user model, a query intent classifier and the maximization of an intent-aware retrieval quality metric. Besides, Andrey claimed that it is possible to build a high-quality vertical search engine for any type of content and presented a way to implement and deploy such verticals elegantly and quickly.

7 Student Support

ECIR has traditionally a strong focus on student participation. This year student support included mentoring program, student grants, and doctoral consortium.

ECIR 2013 mentoring program chaired by Mikhail Ageev (Moscow State University, Russia) was aimed to assist young authors in preparing papers for submission and improve quality of submissions. We received 13 mentoring requests that were assigned to 13 mentors to be supervised for 1.5 months until the paper submission deadline. Two full papers and a short paper authored by mentoring program graduates were accepted for the conference.

Student grant program included fee waivers and accommodation grants. We received 62 grant applications in total, 29 students were granted with free participation and 15 – with free accommodation in the guest house of the HSE. When judging the applications preferences were given to full-time students from developing countries and students presenting their work that have no alternative travel support.

The Doctoral Consortium (DC) took place on Sunday, March 24, 2013, at the Yandex headquarters, and was co-chaired by Hideo Joho (University of Tsukuba, Japan) and Dmitry Ignatov (Higher School of Economics, Russia). We received 9 submissions from the range of countries, and 5 students were invited to the DC to present their work and have feedback from the mentors.

Each of them also had an opportunity to have in-depth discussions with one of the mentors in a separate space.

8 Publicity and Coverage

Besides using traditional channels like popular mailing lists for dissemination of calls for papers, tutorials, workshops and for participation, organizers actively supported two official social network profiles: on Facebook (`https://www.facebook.com/ecir2013`) and on Twitter (`https://twitter.com/ecir2013`). Both accounts announced key steps in preparation of the conference, conference calls, invited speakers, and, during and after the conference, they shared photos and interviews with presenters. These activities resulted into 222 "likes" of ECIR 2013 on Facebook and 322 "followers" of ECIR 2013 on Twitter. During the conference, participants actively shared their impressions on Twitter using #ecir2013 hashtag (`https://twitter.com/search?q=%23ecir2013`) (See Figure 5).

Figure 5: Typical tweets of ECIR participants

All information about the conference was also shared at the official conference web site, featuring a special conference design: `http://ecir2013.org` and all participants received a printed colourful 37-page conference booklet[2], containing conference program, maps, and other useful information, including tips about visiting nearby restaurants and museums.

All keynotes and industry talks were professionally recorded on video with links to them available at the conference web-site[3], [4].Some keynote speakers and best paper award winners have been interviewed "on the spot" and these videos were shared via conference social network profiles.

9 Additional authors

This report and the conference itself would never happen without the concerted effort of the following co-organizers: Eugene Agichtein (Posters Chair), Evgeniy Gabrilovich (Workshop Chair), Djoerd Hiemstra (Tutorial Chair), Dmitry Ignatov (Proceedings Chair), Sergei O. Kuznetsov

[2] `http://download.yandex.ru/company/ECIR_bookletINT.pdf`
[3] `http://ecir2013.org/keynotes.xml`
[4] `http://ecir2013.org/industryday.xml`

(General Co-Chair), Stefan Rueger (PC Co-Chair), Ilya Segalovich (Industry day chair), Arjen de Vries (Best Paper Awards Chair), Emine Yilmaz (Demo Chair).

10 Acknowledgements

We gratefully thank all Program Committee members for their time and efforts ensuring a high level of the ECIR 2013 program. We thank Mikhail Ageev and Hideo Joho and Dmitriy Ignatov for coordinating mentoring program and doctoral consortium, respectively, as well as all mentors volunteered for these activities (Udo Kruschwitz, Benjamin Piwowarski, Ian Ruthven, Fabrizio Silvestri, Manolis Koubarakis, Maarten de Rijke, Allan Hanbury, Massih-Reza Amini, Alvaro Barreiro, Djoerd Hiemstra, Fabio Crestani, Oren Kurland, Patrick Gallinari, Jaap Kamps, Stefan Rüger, Konstantin Vorontsov, Natalia Loukashevitch, Marie-Francine Moens, and Alexander Panchenko). We also thank Dmitriy Ignatov who contributed to the conference as the Proceedings chair.

We would like to thank the invited speakers and Industry Day speakers for a bright assortment of talks.

We appreciate generous financial support from Yandex and HSE, as well as from our sponsors Mail.Ru and Russian Foundation for Basic Research (platinum level), Google and ABBYY (gold level), Yahoo! Labs (silver level), and Microsoft Research (bronze level).

We would also like to express our deep gratitude to Yandex event management team, who helped us through 1.5 years of preparation for the conference and to all volunteers, who came along and made the conference a success.

In this report we would like also to pay the tribute to the memory of Ilya Segalovich, Yandex co-founder and CTO, who passed away in July 2013 at the age of 48. Ilya was not only one of the initiators of holding ECIR in Moscow and industry day chair, but also contributed a lot to the overall conference organization.

References

[1] C. Carpineto and G. Romano. Semantic search log k-anonymization with generalized k-cores of query concept graph. In Serdyukov et al. [5], pages 110–121.

[2] A. Chuklin, P. Serdyukov, and M. de Rijke. Using intent information to model user behavior in diversified search. In Serdyukov et al. [5], pages 1–13.

[3] P. Metrikov, V. Pavlu, and J. A. Aslam. Optimizing nDCG gains by minimizing effect of label inconsistency. In Serdyukov et al. [5], pages 760–763.

[4] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose. Understanding relevance: An fMRI study. In Serdyukov et al. [5], pages 14–25.

[5] P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. M. Rüger, E. Agichtein, I. Segalovich, and E. Yilmaz, editors. *Advances in Information Retrieval – 35th European Conference on IR Research, ECIR 2013, Moscow, Russia, March 24-27, 2013. Proceedings*,

volume 7814 of *Lecture Notes in Computer Science*. Springer, 2013. http://dx.doi.org/10.1007/978-3-642-36973-5.