



35TH ECIR

EUROPEAN

MOSCOW

CONFERENCE

RUSSIA

ON INFORMATION

24 — 27 MARCH

RETRIEVAL

2013

---

ORGANISED BY

**Yandex**



---

PLATINUM SPONSORS



---

GOLD SPONSORS

**Google**

**ABBYY**

---

SILVER SPONSORS

**YAHOO!**

---

BRONZE SPONSORS

Microsoft  
**Research**

---

SUPPORTED BY



Information Retrieval  
Specialist Group



Welcome to ECIR 2013, the 35th European Conference on Information Retrieval.

The event has been jointly organized by Yandex and the National Research University Higher School of Economics (HSE), and supported by the Information Retrieval Specialist Group at the British Computer Society (BCS-IRSG). The conference brings together more than 200 researchers, students, industry leaders and others for four days of presentations, tutorials, workshops, keynotes and, of course, social events. The conference is held in Moscow, Russia — the easternmost location in the history of the ECIR series.

# WELCOME TO MOSCOW!

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% 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. The accepted contributions represent the state of the art in information retrieval, cover a diverse range of topics, propose novel applications, and indicate promising directions for future research. Of the accepted contributions, 66% have a student as the primary author.

We gratefully thank all Program Committee members for their time and efforts ensuring the high quality level of the ECIR 2013 program, and particularly the Program Committee, Poster and Demo chairs: Jaap Kamps (PC Co-Chair), Stefan Ruger (PC Co-Chair), Eugene Agichtein (Poster Chair), Emine Yilmaz (Demo Chair). We also sincerely thank the best paper selection committee chaired by Arjen de Vries. Additionally, ECIR 2013 hosts several tutorials and workshops covering various IR-related topics. We express our gratitude to the Workshop Chair, Evgeniy Gabilovich, and the Tutorial Chair, Djoerd Hiemstra, and the members of their committees.

We would like to thank our invited speakers — Mor Naaman (Rutgers University, Social Media Information Lab) and Diane Kelly, the winner of the Karen Spärck Jones award. The Industry Day takes place on the final day of the conference and features a bright assortment of talks given by prominent researchers and practitioners: Karen Church (Telefónica Research), Paul Ogilvie (LinkedIn), Hilary Mason (bitly), Antonio Gulli (Bing), Alexey Voropaev (Mail.Ru), Jimmy Lin (Twitter/Univ. of Maryland), Marc Najork (Microsoft Research), and Andrey Styskin (Yandex), to whom we express our gratitude. We also greatly appreciate financial support from our sponsors Mail.Ru and the Russian Foundation for Basic Research (platinum level), Google and ABBYY (gold level), Yahoo! Labs (silver level), and Microsoft Research (bronze level).

Добро пожаловать на 35-ю Европейскую конференцию по информационному поиску, добро пожаловать в Москву!

Welcome to the 35th European Conference on Information Retrieval, welcome to Moscow!

Pavel Serdyukov  
Pavel Braslavski  
Sergei O. Kuznetsov  
Dmitry Ignatov  
Ilya Segalovich  
and the rest of the organizers from Yandex and HSE

## RESTAURANT

### 1 GAVROCHE

Timura Frunze st. 11, bld.19/8.

CUISINE: French, Spanish, Italian

CHECK AVERAGE: 1000–1500 rub.

### 2 BELIY JURAVL

Frunzenskaya Embankment, 14/1

CUISINE: Chinese, Korean, Asian

CHECK AVERAGE: 1000–1500 rub.

### 3 JOHN DONNE PUB

L'va Tolstogo st., 18b, BC Stroganov

CUISINE: English, Japanese, wine bar

CHECK AVERAGE: 1000–1500 rub.

### 4 CARABAS

ADDRESS: L'va Tolstogo st., 18b

CUISINE: European, Pan-Asian

CHECK AVERAGE: 1000–1500 rub.

### 5 CORREA'S

Timura Frunze st., 11

CUISINE: European

CHECK AVERAGE: 650–700 rub.

### 6 FORNETTO – BAR&PIZZA

Timur Frunze st., 11, bld. 2

CUISINE: Italian, European

CHECK AVERAGE: 1000–1500 rub.

### 7 CULINARY KARAVAEVI

#### BROTHERS

Timur Frunze st., 11 building 7

CUISINE: European

CHECK AVERAGE : 300–350 rub.

## SIGHTS

### 8 MUSEUM-STATE

#### OF LEO TOLSTOY IN KHAMOVNIKI

L'va Tolstogo st., 21

WORKING HOURS: from 10.00 till 18.00

### 9 TEMPLE OF ST. NIKOLAY

#### OF KHAMOVNIKI –

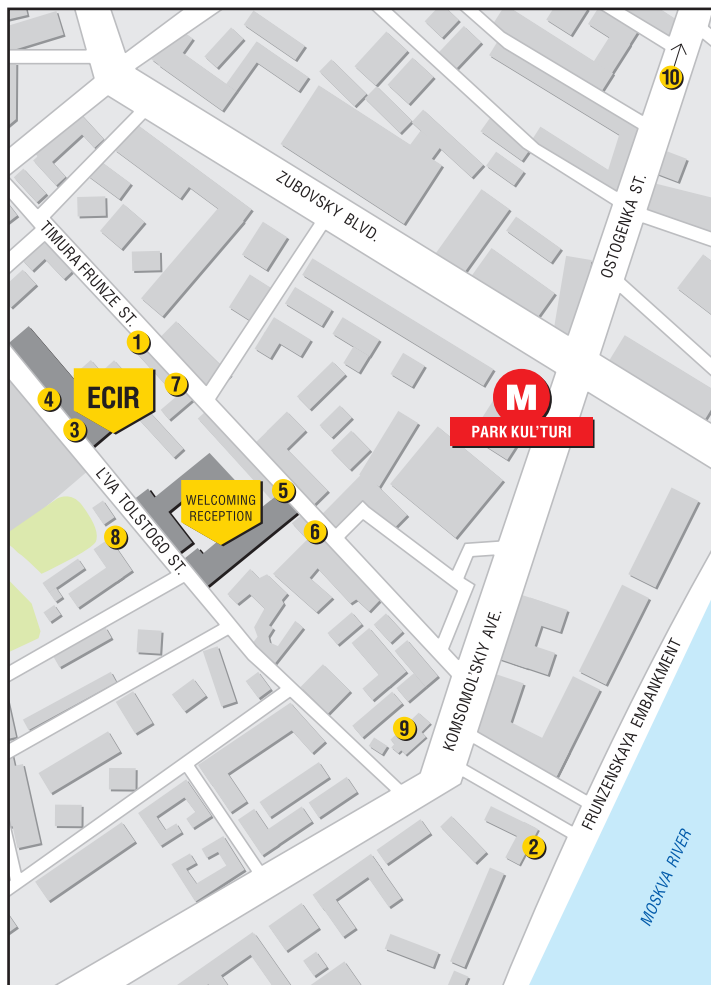
#### CHURCH OF XVIIITH CENTURY.

L'va Tolstogo st., 2

### 10 MULTIMEDIA ART MUSEUM

Ostozhenka St, 16

THE TUTORIAL/WORKSHOPS DAY WILL  
BE HELD AT THE YANDEX  
HEADQUARTERS  
AT LEO TOLSTOY STREET, 18B  
(«STROGANOV» BUILDING)



# PROGRAM

## SUNDAY, 24/MAR/2013

The annual European Conference on Information Retrieval is the main European forum for the presentation of new research results in the field of Information Retrieval. The 35th ECIR conference will take place in Moscow, Russia, from the 24th to the 27th of March 2013. The conference is organised by Yandex, the most popular search engine and most visited website in Russia, in cooperation with one of Russia's best rated and fastest growing universities, the Higher School of Economics, and supported by BCS IRSG.

**09:00 – 10:00** Registration opens

---

**10:00 – 11:30** Tutorials and Workshops

---

**11:30 – 12:00** Coffee break

---

**12:00 – 13:30** Tutorials and Workshops

---

**13:30 – 15:00** Lunch on your own

---

**15:00 – 16:30** Tutorials and Workshops

---

**16:30 – 17:00** Coffee break

---

**17:00 – 19:30** Tutorials and Workshops

---

**20:00 – 22:00** Welcoming reception

---

# TUTORIALS

## SEARCHING THE WEB OF DATA

**GERARD DE MELO** (UC BERKELEY, USA), **KATJA HOSE** (AALBORG UNIVERSITY, DENMARK)

LOCATION: ΟΚΣΦΟΡΔ (OXFORD) 10:00 – 13:30

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.

## DISTRIBUTED INFORMATION RETRIEVAL AND APPLICATIONS

**FABIO CRESTANI AND ILYA MARKOV** (UNIVERSITY OF LUGANO, SWITZERLAND)

LOCATION: ΣΤΕΗΦΟΡΔ (STANFORD) 10:00 – 13:30

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



## TUTORIALS

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.

### **PRACTICAL ONLINE RETRIEVAL EVALUATION**

**FILIP RADLINSKI** (MICROSOFT RESEARCH, UK), **KATJA HOFMANN** (UNIVERSITY OF AMSTERDAM, THE NETHERLANDS)

LOCATION: ОКСФОРД (OXFORD) 15:00 – 19:00

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.

### **CROSS-LINGUAL PROBABILISTIC TOPIC MODELING AND ITS APPLICATIONS IN IR**

**MARIE-FRANCINE MOENS AND IVAN VULIC** (UNIVERSITY OF LEUVEN, BELGIUM)

LOCATION: СТЕНФОРД (STANFORD) 15:00 – 19:00

Cross-lingual topic models are a fairly novel group of unsupervised, language-independent and generative machine learning models that can be effectively trained

## TUTORIALS

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.

# WORKSHOPS

## **INTEGRATING IR TECHNOLOGIES FOR PROFESSIONAL SEARCH**

**MICHAIL SALAMPASIS, NORBERT FUHR, ALLAN HANBURY, MIHAI LUPU, BIRGER LARSEN AND HENRIK STRINDBERG**

LOCATION: СОРБОННА (SORBONNE) FLOOR 2A 10:00 – 19:00

This workshop aims 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.

### **FEATURING KEYNOTES**

Recruiters, Job Seekers and Spammers: Innovations in Job Search at LinkedIn.

**DARIA SOROKINA** (LINKEDIN)

What do we know about how to know about the information needs and behaviors of user groups?

**NICK BELKIN** (RUTGERS UNIVERSITY)

## **FORMAL CONCEPT ANALYSIS (FCA) MEETS INFORMATION RETRIEVAL**

**CLAUDIO CARPINETO, SERGEI O. KUZNETSOV, AMEDEO NAPOLI**

LOCATION: ГАРВАРД (HARVARD) 10:00 – 19:00

Formal Concept Analysis (FCA) is a mathematically well-founded theory aimed at data analysis and classification. 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. In this workshop, we will be interested in two main issues: 1) 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. 2) How can FCA be extended to address a wider range of IR activities, possibly including new retrieval tasks.

### **FEATURING KEYNOTES BY**

**JEAN-FRANÇOIS BOULICAUT** (INSA LYON) **AND BORIS G. MIRKIN** (HSE).

## **DOCTORAL CONSORTIUM**

**HIDEO JOHO AND DMITRIY IGNATOV**

LOCATION: КЕМБРИДЖ (CAMBRIDGE) 10:00 – 19:00

The doctoral students will present summaries of their work to other participating doctoral students and the senior researchers. Each presentation will be followed by a plenary discussion, and individual discussion with one senior advising researcher.

## RESTAURANT

### 1 DOME

Bersenevsky lane, 3/10, bld. 7

CUISINE: European, Italian

CHECK AVERAGE: 1000–1500 rub.

### 2 BONTEMPI

Bersenevskaya Embankment, 12, bld. 1

CUISINE: Italian

CHECK AVERAGE: 1000–1500 rub.

### 3 WUNDERBAR

Bersenevsky lane, 3/10, bld. 7

CUISINE: European

CHECK AVERAGE: 1000–2000 rub.

### 4 PRODUCTY

Bersenevsky lane, 5, bld. 1

CUISINE: Italian

CHECK AVERAGE: 500–1500 rub.

### 5 STRELKA BAR

Bersenevskaya Embankment, 14, bld. 5a

CUISINE: European, international

CHECK AVERAGE: 500–1000 rub.

### 6 RISIRIBA

Serafimovicha st., 2

CHECK AVERAGE: 1000–1500 rub.

### 7 BURO NAKHODOK BAR

Bersenevskaya Embankment, 6

CUISINE: Russian, American

CHECK AVERAGE: 500–1000 rub.

### 8 SCHI SLIVA

Volkhonka st. 9, bld. 1

CUISINE: Russian

CHECK AVERAGE: 1000–1500 rub.

## SIGHTS

### 9 THE PUSHKIN STATE MUSEUM OF FINE ARTS

Volkhonka Street, 12

### 10 THE STATE TRETYAKOV GALLERY

AT KRYMSKY VAL

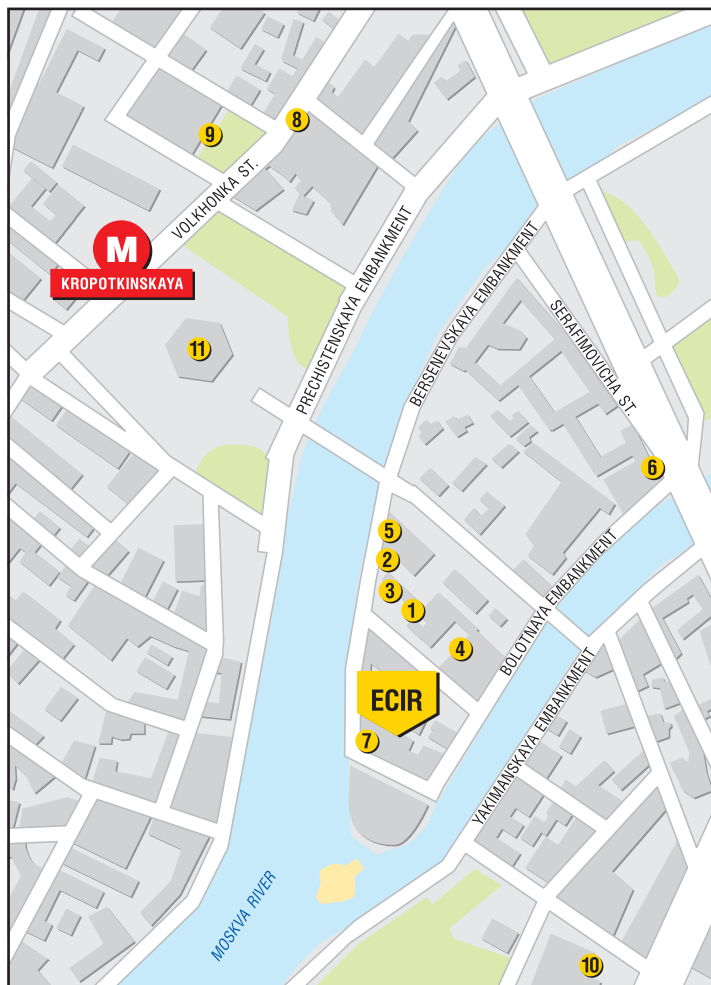
Krymsky Val, 10

### 11 THE CATHEDRAL OF CHRIST

THE SAVIOUR

Volkhonka Street, 15-17

THE MAIN CONFERENCE  
(25 — 27 MARCH, 2013) WILL BE HELD  
AT THE DIGITAL OCTOBER CENTER,  
LOCATED ON BOLOTNY ISLAND  
IN THE MIDDLE OF MOSKVA RIVER  
CONNECTED TO THE MAINLAND  
BY MANY PEDESTRIAN BRIDGES



# PROGRAM

## MONDAY, 25/MAR/2013

9:30 – 10:00

**OPENING**

LOCATION: Conference Hall

10:00 – 11:00

**KEYNOTE BY MOR NAAMAN**

LOCATION: Conference Hall SESSION CHAIR: Jaap Kamps

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

11:00 – 11:30

Coffee break

11:30 – 13:00

**PAPER SESSION 1: USER ASPECTS**

LOCATION: Conference Hall

SESSION CHAIR: Hideo Joho

Using Intent Information to Model User Behavior  
in Diversified Search

**ALEKSANDR CHUKLIN<sup>1,2</sup>, PAVEL SERDYUKOV<sup>1</sup>, MAARTEN DE RIJKE<sup>2</sup>**

1: Yandex, Russia; 2: ISLA, University of Amsterdam

Understanding Relevance: An fMRI Study

**YASHAR MOSHFEGHI, LUISA PINTO, FRANK POLLICK, JOEMON M. JOSE**

University of Glasgow, United Kingdom

An Exploratory Study of Sensemaking  
in Collaborative Information Seeking

**YIHAN TAO, ANASTASIOS TOMBROS**

Queen Mary, University of London, United Kingdom

13.00 – 14.30

Lunch on your own



MONDAY, 25/MAR/2013

14:30 – 16:00

**PAPER SESSION 2A:  
MULTIMEDIA AND CROSS-MEDIA IR**

LOCATION: Conference Hall

SESSION CHAIR: Eric Gaussier

Exploiting User Comments  
for Audio-visual Content  
Indexing and Retrieval

**CARSTEN EICKHOFF<sup>1</sup>, WEN LI<sup>1</sup>,**

**ARJEN P. DE VRIES<sup>2</sup>**

1: Delft University of Technology,

the Netherlands;

2: Centrum Wiskunde &

Informatica, the Netherlands

An Evaluation of Labelling-Game  
Data for Video Retrieval

**RISTE GLIGOROV<sup>1</sup>, MICHIEL HILDEBRAND<sup>1</sup>,**

**JACCO VAN OSSENBRUGGEN<sup>1,2</sup>, LORA AROYO<sup>1</sup>,**

**GUUS SCHREIBER<sup>1</sup>**

1: VU Amsterdam, the Netherlands;

2: CWI Amsterdam, the Netherlands

Multimodal Re-ranking of Product  
Image Search Results

**JOYCE MIRANDA DOS SANTOS, JOÃO MARCOS**

**BASTOS CAVALCANTI, PATRÍCIA CORREIA**

**SARAIVA, EDLENO SILVA DE MOURA**

Federal University of Amazonas (UFAM), Brazi

**PAPER SESSION 2B:  
DATA MINING**

LOCATION: Small Hall

SESSION CHAIR: Patrick Gallinari

An N-Gram Topic Model  
for Time-Stamped Documents

**SHOAB JAMEEL, WAI LAM**

The Chinese University of Hong Kong,

Hong Kong S.A.R. (China)

How Tagging Pragmatics  
Influence Tag Sense Discovery  
in Social Annotation Systems

**THOMAS NIEBLER<sup>1</sup>, PHILIPP SINGER<sup>2</sup>,**

**DOMINIK BENZ<sup>3</sup>, CHRISTIAN KÖRNER<sup>2</sup>,**

**MARKUS STROHMAIER<sup>2</sup>, ANDREAS HOTH<sup>1</sup>**

1: University of Wuerzburg, Germany;

2: Knowledge Management Institute,

Graz University of Technology;

3: Knowledge & Data Engineering Group,

University of Kassel

A Unified Framework for  
Monolingual and Cross-Lingual  
Relevance Modeling Based  
on Probabilistic Topic Models

**IVAN VULIĆ, MARIE-FRANCINE MOENS**

Department of Computer Science, KU Leuven, Belgium

16:00 – 16:30

Coffee break

MONDAY, 25/MAR/2013

PAPER SESSION 3: SHORT PRESENTATIONS

16:30 – 18:30

**PAPER SESSION 3:  
SHORT PRESENTATIONS**

LOCATION: Conference Hall

SESSION CHAIRS: Pavel Braslavski and Pavel Serdyukov

A Versatile Tool for Privacy-Enhanced Web Search

**AVI ARAMPATZIS, GEORGE DROSATOS, PAVLOS S. EFRAIMIDIS**

Democritus University of Thrace, Greece

Exploiting Relevance, Novelty and Diversity  
in Tag Recommendation**FABIANO MUNIZ BELÉM, EDER FERREIRA MARTINS,****MARCOS ANDRÉ GONÇALVES, JUSSARA MARQUES ALMEIDA**

Universidade Federal de Minas Gerais, Brazil

Example Based Entity Finding in the Web of Data

**MARC BRON<sup>1</sup>, KRISZTIAN BALOG<sup>2</sup>, MAARTEN DE RIJKE<sup>1</sup>**1: University of Amsterdam, the Netherlands; 2: Norwegian University of Science  
and Technology, Dept. of Computer and Information ScienceA Fast Generative Spell Corrector based  
on Edit Distance**ISHAN CHATTOPADHYAYA, KANNAPPAN SIRCHABESAN, KRISHANU SEAL**

AOL, India

Being confident about the quality of the predictions  
in Recommender Systems**SERGIO CLEGER-TAMAYO<sup>2</sup>, JUAN M. FERNÁNDEZ-LUNA<sup>1</sup>, JUAN F. HUETE<sup>1</sup>,****NAVA TINTAREV<sup>3</sup>**

1: University of Granada, Spain; 2: University of Holguin, Cuba;

3: University of Aberdeen, UK

Two-Stage Learning to Rank for Information Retrieval

**VAN DANG, MICHAEL BENDERSKY, BRUCE CROFT**

University of Massachusetts Amherst, United States of America



MONDAY, 25/MAR/2013

PAPER SESSION 3: SHORT PRESENTATIONS

16:30 – 18:30

### Hybrid Query Scheduling for a Replicated Search Engine

**ANA FREIRE<sup>1</sup>, CRAIG MACDONALD<sup>2</sup>, NICOLA TONELLOTTI<sup>3</sup>, IADH OUNIS<sup>2</sup>, FIDEL CACHEDA<sup>1</sup>**

1: University of A Coruña, Spain; 2: University of Glasgow, UK;  
3: National Research Council of Italy, Italy

### Latent Factor Blockmodel for Modelling Relational Data

**SHENG GAO<sup>1</sup>, PATRICK GALLINARI<sup>2</sup>**

1: PRIS – Beijing University of Posts and Telecommunications, China, People's Republic of;  
2: LIP6 – Université Pierre et Marie CURIE

### Estimation of the Collection Parameter of Information Models for IR

**PARANTAPA GOSWAMI, ERIC GAUSSIER**

Université Joseph Fourier Grenoble 1, France

### Increasing Stability of Result Organization for Session Search

**DONGYI GUAN, HUI YANG**

Georgetown University, United States of America

### Updating Users About Time Critical News Events

**QI GUO<sup>2</sup>, FERNANDO DIAZ<sup>1</sup>, ELAD YOM-TOV<sup>1</sup>**

1: Microsoft Research; 2: Emory University

### Comparing Crowd-based, Game-based, and Machine-based Approaches in Initial Query and Query Refinement Tasks

**CHRISTOPHER HARRIS, PADMINI SRINIVASAN**

The University of Iowa, United States of America

### Reducing the Uncertainty in Resource Selection

**ILYA MARKOV<sup>1</sup>, LEIF AZZOPARDI<sup>2</sup>, FABIO CRESTANI<sup>1</sup>**

1: University of Lugano, Switzerland; 2: University of Glasgow, UK



MONDAY, 25/MAR/2013

PAPER SESSION 3: SHORT PRESENTATIONS

16:30 – 18:30

Exploiting Time in Automatic Image Tagging

**PHILIP MCPARLANE, JOEMON JOSE**

The University of Glasgow, United Kingdom

Using Text-based Web Image Search Results Clustering  
to Minimize Mobile Devices Wasted Space Interface**JOSE G MORENO, GAËL DIAS**

University of Caen – Lower Normandy, France

Discovery and Analysis of Evolving Topical Social  
Discussions on Unstructured Microblogs**KANIKA NARANG, SEEMA NAGAR, SAMEEP MEHTA, L V SUBRAMANIAM, KUNTAL DE**

IBM Research India, India

Web Credibility: Features Exploration  
and Credibility Prediction**ALEXANDRA OLTEANU, STANISLAV PESHTERLIEV, XIN LIU, KARL ABERER**

EPFL, Switzerland

Query Suggestions for Textual Problem  
Solution Repositories**DEEPAK P<sup>1</sup>, SUTANU CHAKRABORTI<sup>2</sup>, DEEPAK KHEMANI<sup>2</sup>**

1: IBM Research – India, India; 2: Indian Institute of Technology Madras, India

Improving ESA with Document Similarity

**TAMARA POLAJNAR<sup>1</sup>, NITISH AGGARWAL<sup>1</sup>, KARTIK ASSOJA<sup>2</sup>, PAUL BUITELAAR<sup>1</sup>**

1: National University of Ireland, Galway, Ireland;

2: Departamento de Inteligencia Artificial Universidad Politécnica de Madrid

Ontology-Based Word Sense Disambiguation  
for Scientific Literature**ROMAN PROKOFYEV<sup>1</sup>, GIANLUCA DEMARTINI<sup>1</sup>, ALEXEY BOYARSKY<sup>2</sup>,****OLEG RUCHAYSKIY<sup>3</sup>, PHILIPPE CUDRE-MAUROUX<sup>1</sup>**1: University of Fribourg, Switzerland; 2: Instituut-Lorentz for Theoretical Physics, Universiteit  
Leiden, Leiden, The Netherlands; 3: CERN TH-Division, PH-TH, Geneva 23, Switzerland

MONDAY, 25/MAR/2013

PAPER SESSION 3: SHORT PRESENTATIONS

16:30 – 18:30

A Language Modeling Approach for Extracting  
Translation Knowledge from Comparable Corpora

**RAZIEH RAHIMI, AZADEH SHAKERY**

University of Tehran, Islamic Republic of Iran

Content-Based Re-ranking  
of Text-Based Image Search Results

**FRANCK THOLLARD<sup>1</sup>, GEORGES QUÉNOT<sup>2</sup>**

1: LIG – Université Joseph Fourier, France; 2: LIG – CNRS

Encoding Local Binary Descriptors by Bag-of-Features  
with Hamming Distance for Visual Object Categorization

**YU ZHANG, CHAO ZHU, STEPHANE BRES, LIMING CHEN**

liris, France

Recommending High Utility Query via  
Session-Flow Graph

**XIAOFEI ZHU<sup>1,2</sup>, JIAFENG GUO<sup>1</sup>, XUEQI CHENG<sup>1</sup>, YANYAN LAN<sup>1</sup>, WOLFGANG NEJDL<sup>2</sup>**

1: Institute Of Computing Technology, Chinese Academy Of Sciences, Beijing, China;

2: L3S Research Center, Hanover, Germany

URL redirection accounting for improving link-based  
ranking methods

**MAKSIM ZHUKOVSKII, GLEB GUSEV, PAVEL SERDYUKOV**

Yandex, Russian Federation

---

MONDAY, 25/MAR/2013

**18:30 – 20:30 SHORT PAPERS WITH DEMO PRESENTATION**

LOCATION: Demo Room

Re-Leashed! The PuppyIR framework for developing information services for children, adults and dogs

**LEIF AZZOPARDI, DOUG DOWIE**

University of Glasgow, United Kingdom

A Web Mining Tool for Assistance with Creative Writing

**BORIS GALITSKY<sup>1</sup>, SERGEI O KUZNETSOV<sup>2</sup>**

1: eBay, United States of America;

2: Higher School of Economics, Russia

DS4: A Distributed Social and Semantic Search System

**DIONISIS KONTOMINAS<sup>1</sup>,****PARASKEVI RAFTOPOULOU<sup>1</sup>,****CHRISTOS TRYFONOPOULOS<sup>1</sup>,****EURIPIDES G.M. PETRAKIS<sup>2</sup>**

1: University of Peloponnese, Greece;

2: Technical University of Crete, Greece

Serelex: Search and Visualization of Semantically Related Words

**ALEXANDER PANCHENKO<sup>1,2</sup>, PAVEL ROMANOV<sup>2</sup>,****OLGA MOROZOVA<sup>1</sup>, HUBERT NAETS<sup>1</sup>,****ANDREW PHILIPPOVICH<sup>2</sup>, ALEXEY ROMANOV<sup>2</sup>,****CÉDRICK FAIRON<sup>1</sup>**

1: Université catholique de Louvain, Louvain-la-Neuve,

Belgium; 2: Bauman Moscow State Technical University,

Moscow, Russia

**SHORT PAPERS WITH POSTER PRESENTATION**

LOCATION: Location: Poster Room

Lo ajor de dos idiomas — Cross-lingual linkage of geotagged Wikipedia articles

**DIRK AHLERS**

UNITEC, Honduras

A Pilot Study on Using Profile-Based Summarisation for Interactive Search Assistance

**AZHAR HASAN ALHINDI, UDO KRUSCHWITZ,****CHRIS FOX**

University of Essex, United Kingdom

Exploring Patent Passage Retrieval using Nouns Phrases

**LINDA SARA MAUD ANDERSSON<sup>1</sup>, PARVAZ MAHDABI<sup>2</sup>,****ALLAN HANBURY<sup>1</sup>, ANDREAS RAUBER<sup>1</sup>**

1: Vienna University of Technology, Austria;

2: University of Lugano, Switzerland

Investigating the relationship between Usage and Findability within Websites

**LEIF AZZOPARDI, COLIN WILKIE**

University of Glasgow, United Kingdom

Characterizing Health-Related Community Question Answering

**ALEXANDER BELOBORODOV<sup>1</sup>, ARTEM KUZNETSOV<sup>1</sup>,****PAVEL BRASLAVSKI<sup>1,2</sup>**

1: Ural Federal University, Russian Federation; 2: Kontur Labs



SHORT PAPERS WITH DEMO  
PRESENTATION

**18:30 – 20:30** SIAM: Social Interaction Analysis  
for Multimedia  
**JEROME PICAULT, MYRIAM RIBIERE**  
Alcatel-Lucent Bell Labs, France

Exploratory Search  
on Social Media  
**AARON RUSS<sup>1</sup>, MICHAEL KAISER<sup>2</sup>**  
1: German Research Centre for Artificial Intelligence,  
Germany; 2: AGT International, Germany

VisNavi: Citation Context  
Visualization and Navigation  
**FARAG SAAD, BRIGITTE MATHIAK**  
Gesis — ajor z institute for the social sciences, Germany

Face-based People Searching  
in Videos  
**JAN SEDMIDUBSKY, MICHAL BATKO,  
PAVEL ZEZULA**  
Masaryk University, Czech Republic

Political Hashtag Trends  
**INGMAR WEBER<sup>1</sup>, VENKATA RAMA KIRAN  
GARIMELLA<sup>1</sup>, ASMELASH TEKA<sup>2</sup>**  
1: Yahoo! Research, Spain;  
2: Erasmus Mundus master DMKM, Spain

OPARS : Objective Photo  
Aesthetics Ranking System  
**HUANG XIAO, HAN XIAO, CLAUDIA ECKERT**  
Technical University of Munich, Germany

.....

SHORT PAPERS WITH POSTER  
PRESENTATION

A Topic Person Multi-polarization  
Method using Friendship Network  
Analysis  
**ZHONG-YONG CHEN, CHIEN CHIN CHEN**  
National Taiwan University, Taiwan, Republic of China

Improving cyberbullying detection  
with user context  
**MARAL DADVAR<sup>1</sup>, DOLF TRIESCHNIGG<sup>2</sup>, ROELAND  
ORDELMAN<sup>1</sup>, FRANCISKA DE JONG<sup>1</sup>**  
1: Human Media Interaction Group, University of Twente;  
2: Database Group, University of Twente

Snippet-based Relevance  
Predictions for Federated  
Web Search  
**THOMAS DEMEESTER<sup>1</sup>, DONG NGUYEN<sup>2</sup>,  
DOLF TRIESCHNIGG<sup>2</sup>, CHRIS DEVELDER<sup>1</sup>,  
DJOERD HIEMSTRA<sup>2</sup>**  
1: University of Ghent, Belgium;  
2: University of Twente, the Netherlands

Designing Human-Readable User  
Profiles for Search Evaluation  
**CARSTEN EICKHOFF<sup>1</sup>,  
KEVYN COLLINS-THOMPSON<sup>2</sup>, PAUL BENNETT<sup>2</sup>,  
SUSAN DUMAIS<sup>2</sup>**  
1: Delft University of Technology, the Netherlands;  
2: Microsoft Research, United States of America

Sentiment Classification Based  
on Phonetic Characteristics  
**SERGEI ERMAKOV, LIANA ERMAKOVA**  
Perm State National Research University,  
Russian Federation

MONDAY, 25/MAR/2013

SHORT PAPERS WITH POSTER PRESENTATION

18:30 – 20:30

### Cross-Language Plagiarism Detection using a Multilingual Semantic Network

**MARC FRANCO SALVADOR, PARTH GUPTA, PAOLO ROSSO**

Universidad Politécnica de Valencia, Spain

### Classification of Opinion Questions

**HONGPING FU<sup>1</sup>, ZHENDONG NIU<sup>1</sup>, CHUNXIA ZHANG<sup>2</sup>, LU WANG<sup>2</sup>,****PENG JIANG<sup>3</sup>, JI ZHANG<sup>2</sup>**

1: School of Computer Science and Technology, Beijing Institute of Technology, China, People's Republic of; 2: School of Software, Beijing Institute of Technology, China, People's Republic of; 3: HP Labs China

### Tempo of search actions to modeling successful sessions

**KAZUYA FUJIKAWA<sup>1</sup>, HIDEO JOHO<sup>2</sup>, SHIN-ICHI NAKAYAMA<sup>2</sup>**

1: Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan; 2: Faculty of Library, Information and Media Science, University of Tsukuba, Japan

### Near-Duplicate Detection for Online-Shops Owners: an FCA-based Approach

**DMITRY I. IGNATOV, ANDREY V. KONSTANTINOV, YANA CHUBIS**

National Research University Higher School of Economics, Russian Federation

### Incremental Reranking for Hierarchical Text Classification

**QI JU<sup>1</sup>, ALESSANDRO MOSCHITTI<sup>2</sup>**

1: University of Trento, Italy; 2: University of Trento, Italy

### Topic Model for User Reviews with Adaptive Windows

**TAKUYA KONISHI<sup>1</sup>, FUMINORI KIMURA<sup>2</sup>, AKIRA MAEDA<sup>2</sup>**

1: Nara Institute Science and Technology, Japan; 2: Ritsumeikan University, Japan

### Time Based Feedback and Query Expansion for Twitter Search

**NAVEEN KUMAR, BEN CARTERETTE**

University of Delaware, United States of America



MONDAY, 25/MAR/2013

SHORT PAPERS WITH POSTER PRESENTATION

18:30 – 20:30

Is Intent-Aware Expected Reciprocal Rank Sufficient to Evaluate Diversity?

**TEERAPONG LEELANUPAB<sup>1</sup>, GUIDO ZUCCON<sup>2</sup>, JOEMON M. JOSE<sup>3</sup>**

1: King Mongkut's Institute of Technology Ladkrabang, Thailand;

2: Australian e-Health Research Centre, CSIRO, Australia;

3: School of Computing Science University of Glasgow, United Kingdom

Late Data Fusion for Microblog Search

**SHANGSONG LIANG, MAARTEN DE RIJKE, MANOS TSAGKIAS**

University of Amsterdam, The Netherlands

A Task-Specific Query and Document Representation for Medical Records Search

**NUT LIMSOPATHAM, CRAIG MACDONALD, IADH OUNIS**

University of Glasgow, United Kingdom

On CORI Results Merging

**ILYA MARKOV<sup>1</sup>, AVI ARAMPATZIS<sup>2</sup>, FABIO CRESTANI<sup>1</sup>**

1: University of Lugano, Switzerland; 2: Democritus University of Thrace, Greece

Detecting Friday Night Party Photos: Semantics for Tag Recommendation

**PHILIP MCPARLANE<sup>2</sup>, YELENA MEJOVA<sup>1</sup>, INGMAR WEBER<sup>1</sup>**

1: Yahoo! Research, Spain; 2: Glasgow University

Optimizing nDCG Gains by Minimizing Effect of Label Inconsistency

**PAVEL METRIKOV, VIRGIL PAVLU, JAVED ASLAM**

Northeastern University, United States of America

Least squares consensus clustering: criteria, methods, experiments

**BORIS MIRKIN, ANDREY SHESTAKOV**

Highest School of Economics, Russian Federation

MONDAY, 25/MAR/2013

SHORT PAPERS WITH POSTER PRESENTATION

18:30 – 20:30

Domain Adaptation of Statistical Machine Translation models with Monolingual Data for Cross Lingual Information Retrieval

**VASSILINA NIKOULINA, STEPHANE CLINCHANT**

XRCE, France

Text Summarization while Maximizing Multiple Objectives with Lagrangian Relaxation

**MASAAKI NISHINO, NORIHITO YASUDA, TSUTOMU HIRAO, JUN SUZUKI, MASAAKI NAGATA**

NTT Communication Science Laboratories, Japan

Topic Models Can Improve Domain Term Extraction

**MICHAEL NOKEL<sup>1</sup>, ELENA BOLSHAKOVA<sup>1</sup>, NATALIA LOUKACHEVITCH<sup>2</sup>**

1: Moscow State University, Russian Federation;

2: Research Computing Center Moscow State University, Russian Federation

Towards Detection of Child Sexual Abuse Media: Categorization of the Associated Filenames

**ALEXANDER PANCHENKO, RICHARD BEAUFORT, HUBERT NAETS, CÉDRICK FAIRON**

Université catholique de Louvain, Louvain-la-Neuve, Belgium

Leveraging Latent Concepts for Retrieving Relevant Ads For Short Text

**ANKIT PATIL, KUSHAL DAVE, VASUDEVA VARMA**

International Institute of Information Technology Hyderabad, India

Robust PLSA Performs Better Than LDA

**ANNA POTAPENKO<sup>1</sup>, KONSTANTIN VORONTSOV<sup>1,2,3</sup>**

1: Moscow State University, Russian Federation; 2: Moscow Institute of Physics and Technology, Russian Federation; 3: Institution of Russian Academy of Sciences Dorodnicyn Computing Centre, Russian Federation

WANTED: Focused Queries for Focused Retrieval

**GEORGINA RAMIREZ**

Universitat Pompeu Fabra, Spain



MONDAY, 25/MAR/2013

SHORT PAPERS WITH POSTER PRESENTATION

18:30 – 20:30

Exploiting Click Logs for Adaptive Intranet Navigation

**SHARHIDA SAAD, UDO KRUSCHWITZ**

University of Essex, United Kingdom

Leveraging Microblogs for Spatiotemporal Music  
Information Retrieval**MARKUS SCHEDL**

Johannes Kepler University (JKU), Austria

Topic-focused Summarization of Chat Conversations

**ARPIT SOOD, THANVIR P MOHAMED, VASUDEVA VARMA**

IIT Hyderabad, India

Risk Ranking from Financial Reports

**MING-FENG TSAI<sup>1</sup>, CHUAN-JU WANG<sup>2</sup>**

1: National Chengchi University, Taiwan, Republic of China;

2: Taipei Municipal University of Education, Taiwan, Republic of China

Sub-Sentence Extractive Summarization based  
on Combinatorial Optimization**NORIHIITO YASUDA, MASA AKI NISHINO, TSUTOMU HIRAO, MASA AKI NAGATA**

NTT Corporation, Japan

Detecting Expected and Unexpected Adverse  
Drug Reactions from User Reviews  
on Social Media Sites**ANDREW YATES, NAZLI GOHARIAN**

Georgetown University, United States of America

The Impact of Temporal Intent Variability on Diversity  
Evaluation**KE ZHOU<sup>1</sup>, STEWART WHITING<sup>1</sup>, JOEMON JOSE<sup>1</sup>, MOUNIA LALMAS<sup>2</sup>**

1: University of Glasgow, Glasgow, United Kingdom; 2: Yahoo! Lab, Barcelona, Spain



# PROGRAM

## TUESDAY, 26/MAR/2013

10:00 – 11:00

### KEYNOTE BY DIANE KELLY (KAREN SPÄRCK JONES AWARD WINNER)

LOCATION: Conference Hall SESSION CHAIRS: Ayse Goker and John Tait

11:00 – 11:30

Coffee break

11:30 – 13:00

### PAPER SESSION 4: IR THEORY AND FORMAL MODELS

LOCATION: Conference Hall

SESSION CHAIR: Arjen P. de Vries

Semantic Search Log k-Anonymization  
with Generalized k-Cores of Query Concept Graph

**CLAUDIO CARPINETO, GIOVANNI ROMANO**

Fondazione Ugo Bordoni, Italy

A Joint Classification Method to Integrate Scientific  
and Social Networks

**MAHMOOD NESHATI<sup>1</sup>, EHSANEDDIN ASGARI<sup>2</sup>, DJOERD HIEMSTRA<sup>3</sup>, HAMID BEIGY<sup>1</sup>**

1: Sharif University of Technology, Islamic Republic of Iran; 2: Ecole Polytechnique Fédérale de Lausanne – EPFL; 3: Database Research Group, Electrical Engineering, Mathematics and Computer Science (EEMCS) Department, University of Twente

Using Document-Quality Measures to Predict  
Web-Search Effectiveness

**FIANA RAIBER, OREN KURLAND**

Technion – Israel Institute of Technology, Israel

13:00 – 14:30

Lunch on your own



TUESDAY, 26/MAR/2013

**14:30 – 16:00 PAPER SESSION 5A:  
IR SYSTEM ARCHITECTURES**

LOCATION: Conference Hall

SESSION CHAIR: Ingmar Weber

Training Efficient Tree-Based  
Models for Document Ranking

**NIMA ASADI, JIMMY LIN**

University of Maryland, College Park,  
United States of America

DTD-based costs for Tree-Edit  
distance in Structured Information  
Retrieval

**CYRIL LAITANG, KAREN PINEL-SAUVAGNAT,  
MOHAND BOUGHANEM**

IRIT, France

Ranked Accuracy and  
Unstructured Distributed Search

**SAMI RICHARDSON, INGEMAR J. COX**

University College London, United Kingdom

**PAPER SESSION 5B:  
CLASSIFICATION**

LOCATION: Small Hall

SESSION CHAIR: Marie-Francine Moens

Learning to Rank from Structures  
in Hierarchical Text Classification

**QI JU<sup>1</sup>, ALESSANDRO MOSCHITTI<sup>2</sup>, RICHARD  
JOHANSSON<sup>3</sup>**

1: University of Trento, Italy; 2: University of Trento, Italy; 3: University of Gothenburg, Sweden

Folktale classification using  
learning to rank Dong Nguyen,  
Dolf Trieschnigg, Mariet Theune  
Twente University, the Netherlands

Open-set Classification  
for Automated Genre  
Identification

**DIMITRIOS PRITSOS, EFSTATHIOS STAMATATOS**

Aegean University, Greece

.....  
**16:00 – 16:30**

Coffee break  
.....

TUESDAY, 26/MAR/2013

**16:30 – 18:00 PAPER SESSION 6A:  
WEB**

LOCATION: Conference Hall

SESSION CHAIR: Leif Azzopardi

Semantic Tagging of Places  
based on User Interest Profiles**VINOD KUMAR HEGDE, JOSIANE XAVIER PARREIRA,  
MANFRED HAUSWIRTH**

Digital Enterprise Research Institute, Ireland

Sponsored Search Ad Selection  
by Keyword Structure Analysis**KAI HUI<sup>1</sup>, BIN GAO<sup>2</sup>, BEN HE<sup>1</sup>, TIEJIAN LUO<sup>1</sup>**1: University of Chinese Academy of Sciences, China,  
People's Republic of; 2: Microsoft Research AsiaIntent-based browsing activity  
segmentation**YURY USTINOVSKIY, ANNA MAZUR,  
PAVEL SERDYUKOV**

Yandex, Russian Federation

**PAPER SESSION 6B:  
EVENT DETECTION**

LOCATION: Small Hall

SESSION CHAIR: Fabio Crestani

Extracting Event-Related  
Information from Article Updates  
in Wikipedia**MIHAI GEORGESCU, NATTIYA KANHABUA, DANIEL  
KRAUSE, WOLFGANG NEJDL, STEFAN SIERSDORFER**

L3S Research Center / Leibniz Universität Hannover, Germany

Using WordNet Hypernyms  
and Dependency Features  
for Phrasal-level Event Recogni-  
tion and Type Classification**YOONJAE JEONG, SUNG-HYON MYAENG**Korea Advanced Institute of Science and Technology  
(KAIST)Aggregating Evidence from  
Hospital Departments to Improve  
Medical Records Search**NUT LIMSOPATHAM, CRAIG MACDONALD,  
IADH OUNIS**

University of Glasgow, United Kingdom

**19:30 – 23:00**

Conference party



TUESDAY, 26/MAR/2013

## CONFERENCE PARTY WILL BE HELD ON THE EVENING OF TUESDAY, 26TH MARCH, ON A SNOW-WHITE YACHT

From the water, participants will have the unique opportunity to make remarkable pictures of unique quays, the Kremlin, Moscow State University, the Saint Basil's cathedral, the Novodevichy Convent and other world famous monuments amazingly enlightened at night.



# PROGRAM

## WEDNESDAY, 27/MAR/2013

10:00 – 11:30

### INDUSTRY DAY SESSION 1

LOCATION: Conference Hall

SESSION CHAIR: Ilya Segalovich

Search and Discovery at Twitter

**JIMMY LIN**

Research Scientist, Twitter, Associate Professor at University of Maryland

Social Search

**MARC NAJORK**

Principal Researcher, Microsoft Research

11:30 – 12:00

Coffee break

12:00 – 13:30

### INDUSTRY DAY SESSION 2

LOCATION: Conference Hall

SESSION CHAIR: Ilya Segalovich

Lessons from the Wild: How Context Can Shape  
Consumption in Content Recommendation Systems

**PAUL OGILVIE**

Software Engineer, LinkedIn

Opportunities in Web Search

**HILARY MASON**

Chief Scientist, bitly

13:30 – 15:00

Lunch on your own



WEDNESDAY, 27/MAR/2013

**15:00 – 16:30 INDUSTRY DAY SESSION 3**

LOCATION: Conference Hall

SESSION CHAIR: Ilya Segalovich

Suggestion technologies  
for Bing

**ANTONIO GULLI**

Principal Development Manager, Bing Europe

Mobile search: a force  
to be reckoned with!

**KAREN CHURCH**

Researcher, Telefonica Research

**P8B: PAPER SESSION****8B: TEMPORAL IR**

LOCATION: Small Hall

SESSION CHAIR: Fernando Diaz

Predicting Information Diffusion  
in Social Networks  
using Content and User's  
Profiles

**CEDRIC LAGNIER<sup>2</sup>, LUDOVIC DENOYER<sup>1</sup>,****ERIC GAUSSIER<sup>2</sup>, PATRICK GALLINARI<sup>1</sup>**

1: UPMC — LIP6, France; 2: IMAG, France

Influence of Timeline  
and Named-entity Components  
on User Engagement

**YASHAR MOSHFEGHI<sup>1</sup>, MICHAEL MATTHEWS<sup>2</sup>,****ROI BLANCO<sup>2</sup>, JOEMON M. JOSE<sup>1</sup>**

1: University of Glasgow, United Kingdom;

2: Yahoo! Research Lab, Spain

Cognitive Temporal  
Document Priors

**MARIA-HENDRIKE PEETZ, MAARTEN DE RIJKE**

University of Amsterdam, the Netherlands

**16:30 – 17:00**

Coffee break

WEDNESDAY, 27/MAR/2013

**17:00 – 18:30 INDUSTRY DAY SESSION 4**

LOCATION: Conference Hall

SESSION CHAIR: Ilya Segalovich

Active Learning to Rank

**ALEXEY VOROPEV**

Head of Ranking department, Mail.Ru

Aggregate and conquer: Finding  
the way in the diverse world  
of user intents**ANDREY STYSKIN**

Head of Web Ranking Team, Yandex

**P9B: PAPER SESSION****9B: MICROBLOG SEARCH**

LOCATION: Small Hall

SESSION CHAIR: Padmini Srinivasan

Combining Recency  
and Topic-Dependent Temporal  
Variation for Microblog Search**TAIKI MIYANISHI, KAZUHIRO SEKI,****KUNIAKI UEHARA**Graduate School of System Informatics Kobe University,  
JapanSubjectivity annotation  
of the Microblog 2011 Realtime  
Adhoc relevance judgments**GEORGIOS PALTOGLOU, KEVAN BUCKLEY**University of Wolverhampton,  
United KingdomGeo-spatial Event Detection  
in the Twitter Stream**MAXIMILIAN WALTHER, MICHAEL KAISER**

AGT International, Germany

**18:30 – 19:00**

Closing

**PRESENTATION OF ECIR 2014**

LOCATION: Conference Hall

## KEYNOTE SPEAKERS

### MOR NAAMAN

#### **TIME FOR EVENTS: TELLING THE WORLD'S STORIES FROM SOCIAL MEDIA**

ASSISTANT PROFESSOR AT THE RUTGERS SCHOOL OF COMMUNICATION AND INFORMATION



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 Oscar's, 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. Our work tackles 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, analyze, and experience an event through its social media content. Our work 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.



## DIANE KELLY

### **KAREN SPÄRCK JONES AWARD WINNER TALK**

ASSOCIATE PROFESSOR IN THE SCHOOL OF INFORMATION AND LIBRARY SCIENCE AT  
THE UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL



BCS/ BCS IRSG Karen Spärck-Jones Award Panel has decided to make the Award for 2012 to Diane Kelly. In making the BCS/ BCS IRSG Karen Spärck-Jones Award for 2012, the Panel strongly recognizes that Diane has made important contributions to: the analysis of information seeking behaviors, and to the development of new experimental methods and systems to support information seeking and analysis.

Diane has made several other important contributions to user modeling using implicit indicators of relevance, the development and analysis of interfaces to elicit richer statements of interest, and new methodologies for designing and evaluating interactive retrieval systems. Her strong user-oriented work views users-as-people with cognitive tasks.

## INDUSTRY DAY SPEAKERS CHAIR ILYA SEGALOVICH

## JIMMY LIN

RESEARCH SCIENTIST, TWITTER, ASSOCIATE PROFESSOR  
AT UNIVERSITY OF MARYLAND  
SEARCH AND DISCOVERY AT TWITTER



Twitter aims to be an information platform that connects users to what they care about, 140 characters at a time. Whether it's breaking new events around the world, the latest celebrity gossip, or the recent adventures of your closest friends, the search and discovery services aim to surface relevant and personalized content in real-time. In this talk, I will provide an overview of projects at Twitter in this space, discuss challenges that we face, and share some experiences we gained along the way. In particular, I'll focus on the real-time aspects of the problem, which has implications for both the design of search architectures and content-ranking algorithms.

## MARC NAJORK

PRINCIPAL RESEARCHER, MICROSOFT RESEARCH  
SOCIAL SEARCH



This talk discusses the interplay of “Social” and “Search”. Social signals can be used to improve the relevance of algorithmic search results, and information retrieval techniques can be used to surface content from social networks. I will give an overview of the ways in which social content is used to improve the Bing search experience, and at a more abstract level will discuss how signals from a variety of social networks can be used to improve the user's experience in different task scenarios. I will also touch on some problematic aspects of leveraging social information, such as demographics of the user base and trustworthiness of postings.

## PAUL OGILVIE

**SOFTWARE ENGINEER, LINKEDIN**

LESSONS FROM THE WILD: HOW CONTEXT CAN SHAPE CONSUMPTION IN CONTENT RECOMMENDATION SYSTEMS



Cranfield-style evaluations have the advantage of reproducibility but ignore many factors that can shape user interaction. It is common in industry to augment static test collections with tests performed on live user traffic. Split or A/B testing can be used to help control compounding factors, but nevertheless they can have a large impact on online metrics. How users interact with the product also provide a valuable source of data for understanding a product and forming hypotheses. For example, users following an email link messaged “Top content, tailored for you from the people, industries, and companies you’re following” may have a different expectation about the content recommendations than when following a link from the homepage of LinkedIn titled “LinkedIn Today recommends this content for you.” In this talk, we share lessons and observations about patterns of user interaction with the content recommendations we provide in LinkedIn Today. We pay special attention to how members arrive at the product, the different contexts in which the product may be consumed (website, mobile, email), and what they do next.

## HILARY MASON

**CHIEF SCIENTIST, BITLY**

OPPORTUNITIES IN WEB SEARCH



Search is not a solved problem! We still lack robust products, algorithms, and tools for non-query driven web search. This talk explores progress in search of social data and real-time data, and looks at practical methods for query large volumes of social data in a real-time system; why this can help improve the quality of search and user experience.

## INDUSTRY DAY SPEAKERS

### ANTONIO GULLI

**PRINCIPAL DEVELOPEMENT MANAGER, BING EUROPE**

SUGGESTION TECHNOLOGIES FOR BING



Search engines have long provided clues to the topics people look up. In this talk, we discuss Bing suggestion technologies (autosuggest) for the use of showing the precise questions that are most frequently asked. These technologies anticipate what you are most likely to ask based on questions that other people have made. Simply type a question starting with a word like “is” or “was,” and search engines will start filling in the rest. Frequently asked questions include: “When will the world end?” “Is Neil Armstrong Muslim?” “What is the meaning of the life”. During the talk, we will demonstrate why autosuggest reflects the collective curiosities of its users.

### KAREN CHURCH

**RESEARCHER, TELEFONICA RESEARCH**

MOBILE SEARCH: A FORCE TO BE RECKONED WITH!



Recently the world has witnessed a revolution in terms of mobile web and mobile search usage. Mobile phones, once deemed as simple communications devices, now provide mobile users with access to a wealth of online content, anytime and anywhere. In 2012, the increasing presence of mobile devices caused desktop search to decline for the first time ever; a level of growth that simply cannot be ignored. In this talk, I'll take a nostalgic look back at the simple beginnings of mobile search and discuss how, why and in what ways mobile search has evolved over the past eight years. I'll highlight patterns of mobile search usage and show how they not only differ from desktop search, but they are continually evolving. And instead of taking a single, data-centric viewpoint of mobile search, I'll also discuss user-centric studies, highlighting the unique needs, intents and motivations of mobile searchers. Finally, I'll share some of my thoughts about where mobile search is heading, the challenges that lie ahead and discuss some of the factors that I think are important when it comes to enriching the future search experiences of mobile users.

## ALEXEY VOROPAEV

**HEAD OF RANKING DEPARTMENT, MAIL.RU**  
ACTIVE LEARNING TO RANK



Development of a system based on supervised machine learning includes three main steps: factors selection, building training set and appropriate ML algorithm application. The training set construction is the very problematic aspect, since usually it is not well controlled but may dramatically affect the resulting quality of ML model. In my talk I am going to introduce our active learning technique to manipulate the training set in the context of learning to rank problem. Using simple and effective algorithm we can significantly reduce the training set size as well as improve the ranking quality.

## ANDREY STYSKIN

**HEAD OF WEB RANKING TEAM, YANDEX**  
AGGREGATE AND CEONQUER: FINDING THE WAY IN THE DIVERSE WORLD OF USER INTENTS



In our days a search engine result page is not represented by just 10 blue links. It usually contains additional special result items, containing news, image or video results. These verticals are intended to satisfy specific user intents, but those needs are actually very diverse and search engines are often not able to provide, for example, a special vertical search engine for some rare user need like “buy a 3-wheel bicycle”. In this talk I am going to describe the approach adopted by Yandex. Instead of fixing slots for vertical documents, we adopt a more flexible technique based on a probabilistic user model. It allows us to deal with redundant and complimentary verticals in the situations when not all verticals are equally relevant and important for the query. Using our framework we can easily create a new vertical ranking using the existing web document collection. This ranking can then be easily incorporated into the aggregated search results page to maximize the chance of satisfying the user.

