19th International Conference on Information and Knowledge Management

Toronto, Canada

October 26 - 30, 2010

CIKM Toronto 2010

THE Fairmont ROYAL YORK

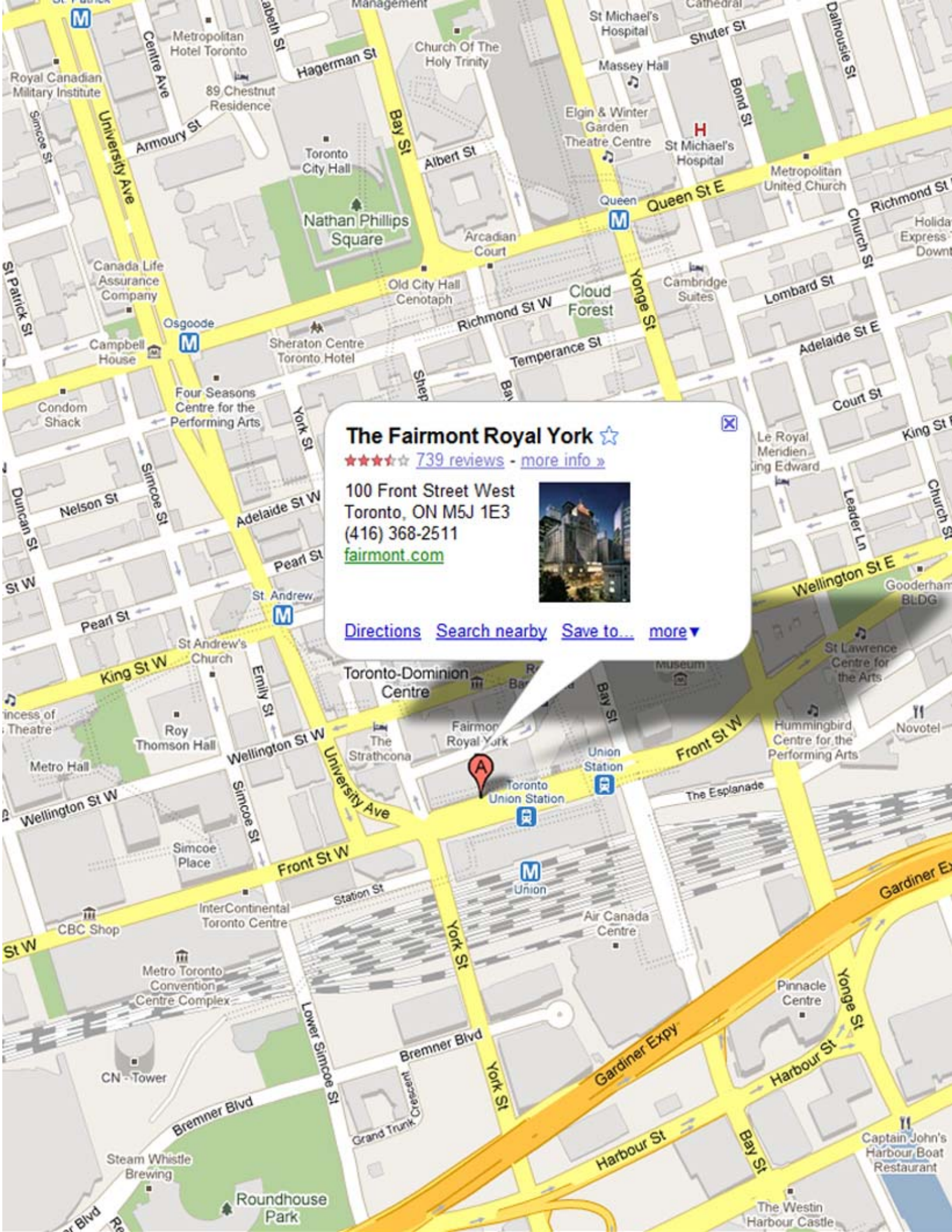

The Fairmont Royal York
739 reviews - more info »
100 Front Street West
Toronto, ON M5J 1E3
(416) 368-2511
fairmont.com
Directions
Search nearby
Save to...
more
Metropolitan Hotel Toronto
Royal Canadian Military Institute
89 Chestnut Residence
Church Of The Holy Trinity
St Michael's Hospital
Massey Hall
Elgin & Winter Garden Theatre Centre
Toronto City Hall
Nathan Phillips Square
Arcadian Court
Metropolitan United Church
Canada Life Assurance Company
Old City Hall Cenotaph
Cloud Forest
Cambridge Suites
Osgoode
Campbell House
Sheraton Centre Toronto Hotel
Four Seasons Centre for the Performing Arts
Condom Shack
Le Royal Meridien King Edward
St. Andrew
St Andrew's Church
Toronto-Dominion Centre
Roy Thomson Hall
Metro Hall
The Strathcona
Fairmont Royal York
Union Station
Toronto Union Station
Hummingbird Centre for the Performing Arts
St Lawrence Centre for the Arts
Novotel
Simcoe Place
InterContinental Toronto Centre
CBC Shop
Metro Toronto Convention Centre Complex
Air Canada Centre
Pinnacle Centre
CN - Tower
Steam Whistle Brewing
Roundhouse Park
Captain John's Harbour Boat Restaurant
The Westin Harbour Castle
Queen
Queen St E
Richmond St W
Richmond St E
Wellington St E
Wellington St W
Front St W
Adelaide St W
Adelaide St E
Bay St
Yonge St
University Ave
Simcoe St
Lower Simcoe St
York St
Bremner Blvd
Harbour St
Gardiner Expy
Gardiner Ex
Lombard St
Church St
Dalhousie St
Bond St
Shuter St
Albert St
Hagerman St
Armoury St
Centre Ave
Nelson St
Pearl St
King St W
Emily St
Station St
Grand Trunk Crescent
Leader Ln
Court St
Temperance St
Duncan St
St Patrick St
Simcoe St
Gooderham BLDG

Table of Contents

Message from the Conference Chair 2
Message from the Program Committee Chairs 4
CIKM'10 Conference Organization 5
CIKM'10 Program at a Glance 6
CIKM'10 Keynote Talks 7
Search Engine Support for Software Applications 7
Schema Extraction 8
Use of Semantics in Real Life Applications 9
Temporal Dynamics and Information Retrieval 10
Industry Track Invited Talks 11
The Shift from Search to Matching 11
Turing Dark Clouds into Silver Linings 12
On-line Social Networks: Insight, Commercial Value and Computational Challenges 13
Digital Advertising: Personalization, Personalization, and Personalization 14
CIKM'10 Main Conference Program 16
Wednesday, October 27, 2010 16
Thursday, October 28, 2010 20
Friday, October 29, 2010 24
CIKM'10 Posters 27
DB Track 27
IR Track 29
KM Track 32
Industry Track 37
CIKM'10 Workshops and Tutorials at a Glance 38
Workshop Programs 39
Fourth Workshop on Analytics for Noisy Unstructured Text Data (AND 2010) 39
Third Workshop on Research Advances in Large Digital Book Repositories and Complementary Media (BooksOnline 2010) 41
ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO 2010) 43
Third International Workshop on Patent Information Retrieval (PaIR 2010) 45
Second International Workshop on Cloud Data Management (CloudDB 2010) 47
ACM Thirteenth International Workshop on Data Warehousing and OLAP (DOLAP 2010) 49
Third Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR 2010) 51
ACM Workshop for Ph.D. Students in Information and Knowledge Management (PIKM 2010) 53
Second International Workshop on Search and Mining User-Generated Contents (SMUC 2010) 55
CIKM'10 Tutorials 57
Tutorial 1: Harvesting Knowledge from Web Data and Text 57
Tutorial 2: Online Advertising Business Models, Technologies and Issues 58
Tutorial 3: Real World Text Mining 59
Tutorial 4: Mobility Data: Modeling, Management, and Understanding 60
Restaurant Information 61
Liberty Grand Banquet Dinner Information 62
Royal York Hotel Maps 63

Message from the Conference Chair

On behalf of the organizing committee, I wholeheartedly welcome you to the 19th ACM International Conference on Information and Knowledge Management (CIKM 2010). I hope this conference proves to be interesting and beneficial.

CIKM is a well-known top tier and premier ACM conference in the areas of information retrieval, knowledge management and database. Since its inception, the CIKM conference has provided a unique international forum for the presentation, discussion, and dissemination of research findings in data management, information retrieval, and knowledge management. The purpose of the conference is to identify challenging problems facing the development of future knowledge and information systems, and to shape future research directions through the publication of high quality, applied and theoretical research findings. The conference has been a leading forum in which experts from academia, industry, and the government gather to exchange ideas, research achievements, and technical developments in multidisciplinary research areas.

CIKM has rapidly grown to become one of the world's most recognized conferences in the field. This year CIKM has received a record high number of submissions in the history of CIKM, as can be seen from the following statistics:

- 1382 abstracts submitted
- 945 full papers plus 38 demo papers submitted
- 126 papers accepted for presentation as full papers *(13.3% acceptance rate)* and an additional 165 were accepted for short papers *(17.5%)*

In addition to regular research tracks, CIKM 2010 features 4 keynote speakers, 4 pre-conference tutorials, 9 workshops, 12 industrial full papers and 21 demo papers. I am proud of our program and acknowledge the tireless efforts of people who materialized this program.

First of all, I am honored to have 4 distinguished keynote speakers: Jamie Callan, Susan Dumais, Gregory Grefenstette, and Divesh Srivastava. I deeply appreciate their time and commitment to deliver their speeches and share their cutting-edge research experiences and insightful comments in their research topics.

An enormous amount of effort was undertaken to materialize this program and produce the proceedings. The technical program of the conference was selected by a renowned program committee that consists of three PC co-chairs, Gareth Jones, Nick Koudas and Xindong Wu, 38 senior program committee members who managed the program's subareas, and 379 program committee members. They were challenged in selecting excellent and high quality papers from many good papers. I wish to express our deepest gratitude to the program committee members, external reviewers, and all authors who submitted their papers to CIKM 2010, presenters and delegates for their contributions and participation. I wish to convey our sincere thanks to ACM for providing professional help and support in compiling the proceedings. I especially acknowledge the contribution of Aijun An who monitored and managed the editing of the proceedings to meet the production schedule and create the program.

I would also like to thank: Industry Track Chair, Kevyn Collins-Thompson; Proceedings Chair: Aijun An; Workshop Chair, Mounia Lalmas; Tutorial co-Chairs, Andrew MacFarlane and Dimitri

Theodoratos; Demo co-Chairs, Udo Kruschwitz and Cathal Gurrin; Local Arrangement Co-Chairs, Xiaohui Yu and Marshall Walker; Sponsorship Chair, Tapas Kanungo; Publicity co-Chairs, Jing Bai and Shusaku Tsumoto; Travel Award co-Chairs, Xiaohua (Tony) Hu and Min Song; Registration Chair, Vivian Hu; Treasurer, E.K. Park; CIKM 2010 Best Interdisciplinary Paper Award Committee members; and Webmasters, Dawid Kasperowicz and Jun Miao. Without their tireless work, this conference could not have been successful.

In addition, I would like to express our appreciation to all of our sponsors: ACM, ACM SIGIR, ACM SIGWEB, SIGKDD, Microsoft Research, eBay, SoSo, Yahoo!, Nokia, Google, Infobright, IRF, MITACS, Yandex, UMBC, Church and Duncan Group Inc., IBM, DossierView, Information Retrieval and Knowledge Management Research Lab and York University. Our special thanks go to Microsoft and MITACS for sponsoring the graduate students travel award and Google for sponsoring the "Best Interdisciplinary Paper" award. Their generous sponsorships and support made this conference successful.

It is impossible to organize a high quality conference without the enormous support from various individuals. Our sincere thanks go to the conference Web support team and ConfMaster for their wonderful contributions. In particular, I would like to acknowledge the generous help received from Vivian Hu, Jessie Zhao, Dawid Kasperowicz, Jun Miao, Jianhan Zhu, Parke Godfrey, Ellis Lau, Kay Li, Xiaofeng Zhou, Zheng Ye, Grace Zhou, Cathy Hu, Chenghua Li, Mariam Daoud, Shirley Hu, Mehdi Kargar, Ameeta Agrawal, Bahareh Sarrafzadeh and David Wallace, who are helping with local arrangements, and providing registration and financial management. Our special thanks go to the Offices of the President, Vice-President Academic & Provost, the Vice-President Research & Innovation and the Dean's Office of the Faculty of Liberal Arts & Professional Studies, York University.

Last, but not least, CIKM 2010 owes a lot to the following who have been nothing short of fantastic in their time and guidance: E.K Park, David Cheung, Charles Nicholas, Il Yeol Song, Ophir Frieder, Ricardo Baeza-Yates, and Nick Cercone.

Finally, I would like to thank the CIKM Steering Committee for their vote of confidence in us. It is a great honor and pleasure to accept the responsibilities and challenges of general chair. I am pleased to offer this excellent program. I hope that attendees find the technical program of CIKM 2010 to be interesting and productive towards their research endeavors. I hope you all take advantage of these opportunities for professional development. I further hope that the conference will be stimulating, informative, enjoyable, and a fulfilling experience to all who attend it.

Enjoy the CIKM 2010 conference and your stay in Toronto, Canada!

Jimmy Huang
ACM CIKM 2010 General Chair
York University, Canada

Message from the Program Committee Chairs

It is our great pleasure to welcome you to the 19th ACM Conference on Information and Knowledge Management (CIKM 2010) in Toronto, Canada. Since 1992, CIKM has been successfully bringing together leading researchers from the database, information retrieval, and knowledge management communities. The purpose of CIKM is to identify challenging problems facing the development of future knowledge and information systems, and to shape future research directions through the publication of high-quality research findings, both theoretical and applied. CIKM 2010 aims to continue this tradition of promoting these aims via exchange and collaboration between the general areas of databases, information retrieval, and knowledge management. Following the success of CIKM 2009 in Hong Kong, this year's call for papers attracted a new record number of 945 full paper submissions plus 38 demo paper submissions from across the world. All submitted papers were reviewed by at least three program committee members, and all the conflictive cases were further reviewed and resolved by one senior program committee member and one program chair. There were a few boarder-line papers, which were re-reviewed by at least one additional program committee member and the program co-chairs. Following a rigorous reviewing process, the program committee accepted 126 submissions as full papers for oral presentation *(13.3%)* and an additional 165 as short papers for poster presentation *(17.5%)*.

We would like to acknowledge everyone who made this technical program possible. First, we would like to thank the authors of all submissions to the conference for providing the contents of the program: the conference would not be possible without your contributions. We express our sincere gratitude to the members of the senior program committees, the regular program committee members and the additional reviewers, who worked extremely hard to provide comments on all submissions within an extremely tight timeframe. Their work was invaluable in forming the final program and providing feedback to the authors. Also, a very big "thank you" goes to the many others who contributed to making CIKM 2010 a success. Finally, we would like to thank all our corporate sponsors for their generous support and Professor Jimmy Huang, the Conference General Chair, for his coordination, guidance, and tremendous efforts in running the different tracks as a unified conference.

We hope that you will find the CIKM 2010 program interesting, thought-provoking and inspiring. Please enjoy the conference and the opportunity to network with friends and colleagues from around the world. Finally, see you next year in Glasgow, Scotland!

Gareth Jones
CIKM 2010 IR Chair
Dublin City University, Ireland

Nick Koudas
CIKM 2010 DB Chair
University of Toronto, Canada

Xindong Wu
CIKM 2010 KM Chair
University of Vermont, USA

CIKM'10 Conference Organization

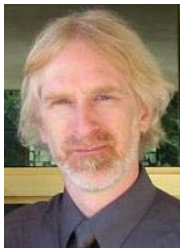
General Conference Chair:	Jimmy Huang *(York University, Canada)*
DB Track Program Chair:	Nick Koudas *(University of Toronto, Canada)*
IR Track Program Chair:	Gareth Jones *(Dublin City University, Ireland)*
KM Track Program Chair:	Xindong Wu *(University of Vermont, USA)*
Industry Track Chair:	Kevyn Collins-Thompson *(Microsoft Research, USA)*
Proceedings Chair:	Aijun An *(York University, Canada)*
Tutorials Co-Chairs:	Andrew MacFarlane *(City University in London, UK)*
	Dimitri Theodoratos *(New Jersey Institute of Tech., USA)*
Workshops Chair:	Mounia Lalmas *(University of Glasgow, UK)*
Demo Co-Chairs:	Cathal Gurrin *(Dublin City University, Ireland)*
	Udo Kruschwitz *(University of Essex, UK)*
Local Arrangment Co-Chairs:	Marshall Walker *(York University, Canada)*
	Xiaohui Yu *(York University, Canada)*
Sponsorship Chair:	Tapas Kanungo *(Microsoft Bing R&D, USA)*
Publicity Co-Chairs:	Jing Bai *(Microsoft Bing R&D, USA)*
	Shusaku Tsumoto *(Shimane University, Japan)*
Travel Award Co-Chairs:	Xiaohua Tony Hu *(Drexel University, USA)*
	Min Song *(New Jersey Institute of Technology, USA)*
Treasurer:	E.K. Park *(City University of New York, USA)*
Steering Committee Chair:	E.K. Park *(City University of New York, USA)*
Steering Committee:	Ophir Frieder *(Georgetown Univeristy, USA)*
	David Grossman *(Illinois Institute of Technology, USA)*
	Albert Larnder *(Universidade Federal de Minas Gerais, Brazil)*
	Il Yeol Song *(Drexel University, USA)*
	Philip S. Yu *(University of Illinois at Chicago, USA)*
Special Advisory Committee:	Ricardo Baeza-Yates *(Yahoo! Research, Spain/Chile)*
	David Cheung *(University of Hong Kong, Hong Kong)*
	Francisco M. Couto *(Universidade de Lisboa, Portugal)*
	Alberto A. F. Laender *(Universidade Federal de Minas Gerais, Brazil)*
	Andre Falcao *(Universidade de Lisboa, Portugal)*
	Pedro Fernandes *(Instituto Gulbenkian de Ciencia, Portugal)*
	Dongwon Lee *(Penn State university, USA)*
	Jan Pedersen *(Microsoft, USA)*
	James G. Shanahan *(Church and Duncan Group Inc., USA)*
	Mario J. Silva *(Universidade de Lisboa, Portugal)*
	Il-Yeol Song *(Drexel University, USA)*
CIKM 2011 Liaison:	Iadh Ounis *(University of Glasgow, UK)*
Registration Chair:	Vivian Hu *(York University, Canada)*
Poster Co-Chairs:	Chenghua Li *(York University, Canada)*
	Jessie Zhao *(York University, Canada)*
Conference Web Masster:	Dawid Kasperowicz *(York University, Canada)*
	Jun Miao *(York University, Canada)*

CIKM'10 Program at a Glance

On-Site Registration: October 26-30, 7:45-18:00 *[Second Floor Lobby]*					
Tuesday, October 26: ***Workshops and Tutorials***					
Wednesday October 27, 2010	**8:10-8:30:** ***Conference Opening*** *[Imperial Room]*				
	8:30-9:30: ***Keynote Talk*** *[Imperial Room]* **Search Engine Support for Software Applications, Jamie Callan, Carnegie Mellon University**				
	9:30-10:00: ***Coffee Brreak*** *[Imperial Foyer]*				
	10:00 - 12:00	**Session 1A** *[Room: Alberta]* ***KM: Information Extractions***	**Session 1B** *[Room: Quebec]* ***IR: Ranking and Retrieval Model***	**Session 1C** *[Room: British Columbia]* ***DB: Indexes and Query Optimization***	**Session 1D** *[Room: Tudor 7 & 8]* ***IR: Domain-Specific and Multimedia IR***
	12:00-13:30: ***Lunch and Posters*** *[Imperial Room]*				
	13:00 - 15:30	**Session 2A** *[Room: Alberta]* ***KM: Link and Graph Mining***	**Session 2B** *[Room: Quebec]* ***IR: Machine Learning for IR (I)***	**Session 2C** *[Room: British Columbia]* ***DB: Mobile and Distributed Data***	**Session 2D** *[Room: Tudor 7 & 8]* ***KM: Classification and Clustering***
	15:30-16:00: ***Coffee Break*** *[British Columbia Foyer]*				
	16:00 - 17:40	**Session 3A** *[Room: Alberta]* ***KM: Large-Scale Statistical Techniques***	**Session 3B** *[Room: Quebec]* ***IR: Machine Learning for IR (II)***	**Session 3C** *[Room: British Columbia]* ***DB: Top-K and Shortest Path Processing***	**Session 3D** *[Room: Tudor 7 & 8]* ***IR: IR Evaluation***
	18:00-21:00: ***Conference Reception and Posters*** *[Imperial Room]*				
Thursday October 28, 2010	**8:30-9:30:** ***Keynote Talk*** *[Imperial Room]* **Schema Extraction, Divesh Srivastava, AT&T Labs-Research**				
	9:30-10:00: ***Coffee Break*** *[Imperial Foyer]*				
	10:00 - 12:00	**Session 4A** *[Room: Alberta]* ***KM: Information Filtering and Recommender Systems (I)***	**Session 4B** *[Room: Quebec]* ***IR: Social Networks and Text Mining***	**Session 4C** *[Room: British Columbia]* ***Industry: IR Applications***	**Session 4D** *[Room: Tudor 7 & 8]* ***DB: Information Retrieval in Databases***
	12:30-13:30: ***Lunch*** *[On Your Own]*				
	13:30 - 15:30	**Session 5A** *[Room: Alberta]* ***KM: Temporal, Spatial and Stream Data Mining***	**Session 5B** *[Room: Quebec]* ***IR: Filtering and Recommendation***	**Session 5C** *[Room: British Columbia]* ***Industry: Databases and OLAP***	**Session 5D** *[Room: Tudor 7 & 8]* ***KM: Data Pre- and Post- Processing***
	15:30-16:00: ***Coffee Break*** *[Imperial Foyer]*				
	16:00 - 17:30	**Demo Session A** *[Room: Imperial]* ***IR(I)***	**Demo Session B** *[Room: Imperial]* ***IR(II)***	**Demo Session C** *[Room: Imperial]* ***DB***	**Demo Session D** *[Room: Imperial]* ***KM***
	18:30-21:30: ***Conference Banquet*** *[Governor's Room, Liberty Grand, 25 British Columbia Road]*				
Friday October 29, 2010	**8:30-9:30:** ***Keynote Talk*** *[Imperial Room]* **Use of Semantics in Real Life Applications, Gregory Grefenstette, Exalead**				
	9:30-10:00: ***Coffee Break*** *[Imperial Foyer]*				
	10:00 - 12:00	**Session 6A** *[Room: Alberta]* ***KM: Information Filtering and Recommender Systems (II)***	**Session 6B** *[Room: Quebec]* ***IR: User Modeling and Search Personalization***	**Session 6C** *[Room: British Columbia]* ***Industry: Web and Social Networks***	**Session 6D** *[Room: Tudor 7 & 8]* ***IR: Query Analysis and Feedback***
	12:00-13:30: ***Lunch*** *[Ballroom]*				
	13:30-14:30: ***Keynote Talk*** *[Imperial Room]* **Temporal Dynamics and Information Retrieval, Susan Dumais, Microsoft Research**				
	14:30-15:00: ***Coffee Break*** *[Imperial Foyer]*				
	15:00 - 17:00	**Session 7A** *[Room: Alberta]* ***KM: Semantic Techniques***	**Session 7B** *[Room: Quebec]* ***IR: Web Search***	**Session 7C** *[Room: British Columbia]* ***Industry: Text Mining and Analytics***	**Session 7D** *[Room: Tudor 7 & 8]* ***IR: Scalability and Efficiency***
Saturday, October 30, 2010: ***Workshops***					

CIKM'10 Keynote Talks

Search Engine Support for Software Applications



Prof. Jamie Callan
Carnegie Mellon University, USA
http://www.cs.cmu.edu/~callan/

Time: October 27, 8:30-9:30
Location: Imperial Room

Abstract

Question-answering, computer-assisted language learning, text mining, and other software applications that use a full-search engine to find information in a large text corpus are becoming common. A software application may use metadata and text annotations to reduce the mismatch between the concept-based representations convenient for inference and the word-based representations typically used for text retrieval. Software applications may also be able to specify detailed requirements that retrieved passages must satisfy. This use of text search is very different than the ad-hoc, interactive search that information retrieval research typically studies.

Search engine developers are beginning to respond by extending indexing and retrieval models developed for structured (e.g., XML) documents to support multiple representations of document content, text annotations, metadata, and relationships. These new requirements force developers to reconsider basic assumptions about index data structures and ranked retrieval models.

How best to use these new capabilities is an open problem. Straightforward transformation of a detailed information need into a complex structured query can produce a query that is effective for exact-match retrieval, but a challenge for the retrieval model to use effectively for best-match retrieval. Bag-of-words retrieval is often disparaged, but its advantage is that it is robust: It works well even when desired documents do not exactly meet expectations.

This talk discusses some of the problems encountered when extending a search engine to support queries posed by other software applications and structured documents with derived annotations.

Biography

Jamie Callan is a Professor of Computer Science at Carnegie Mellon's Language Technologies Institute and School of Information Systems & Management. Prior to joining CMU he was a Research Assistant Professor at the University of Massachusetts, where he also received his Ph.D. His research and teaching focus on text-based information retrieval, primarily search engine architectures, federated search of groups of search engines, adaptive information filtering, text mining, and information retrieval for educational applications. He has published more than 150 scientific papers. He is the Editor-in-Chief of ACM Transactions on Information Systems (TOIS) and was a founding Editor-in-Chief of Foundations and Trends in Information Retrieval (FnTIR). He has served as Chair of ACM SIGIR, and Program Chair of the SIGIR and CIKM conferences.

Schema Extraction

Dr. Divesh Srivastava
AT&T Labs-Research, USA
http://www2.research.att.com/~divesh/

Time: October 28, 8:30-9:30
Location: Imperial Room

Abstract

Understanding the schema of a complex database is a crucial step in exploratory data analysis. However, gaining such an understanding is challenging for new users for many reasons.

First, complex databases often have thousands of inter-linked tables, with little indication of the important tables or the main concepts in the database schema.

Second, schemas can be inaccurate, e.g., some foreign/primary key relationships are not known to designers but are inherent in the data, while others become invalid due to data inconsistencies. In this talk, we present an approach to effectively address these challenges and automatically extract an understandable schema from a complex database.

The first step in our approach is a robust algorithm to discover foreign/primary key relationships between tables. We present a general rule, termed Randomness, that subsumes a variety of other rules proposed in previous work, and develop efficient approximation algorithms for evaluating randomness, using only two passes over the data.

The second step is a principled approach to summarize the schema consisting of tables linked using foreign/primary keys, so that a user can easily identify the main concepts and important tables. We present an information theoretic approach to identify important tables, and an intuitive notion of table similarity that can be used to cluster tables into the main concepts of the schema. We validate our approach using real and synthetic datasets.

This is based on joint work with Marios Hadjieleftheriou, Beng Chin Ooi, Cecilia M. Procopiuc, Xiaoyan Yang and Meihui Zhang.

Biography

Divesh Srivastava is the head of the Database Research Department at AT&T Labs-Research. He received his Ph.D. from the University of Wisconsin, Madison, and his B.Tech from the Indian Institute of Technology, Bombay. He is on the board of trustees of the VLDB Endowment, the associate Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering, and an associate editor of the ACM Transactions on Database Systems. He has served as the program committee co-chair of many conferences, including VLDB 2007. His research interests and publications span a variety of topics in data management.

Use of Semantics in Real Life Applications

Dr. Gregory Grefenstette
Exalead
http://labs.exalead.com/

Time: October 29, 8:30-9:30
Location: Imperial Room

Abstract

Much research in computer exploitable semantics performed in the 1990s and early 2000s is finally finding its way into practical applications. This talk will explore these paths from research to industry, illustrated by current products on the market.

Biography

Gregory Grefenstette (Stanford'78, University of Pittsburgh'93) is chief science officer of Exalead, part of DassaultSystemes. An expert in natural language processing, Grefenstette established the field of Cross Language Information Retrieval by organizing its first Workshop at SIGIR'96. He is also one of the pioneers of distributional semantics, following his PhD work "Exploring Automatic Thesaurus Generation" (Kluwer, 1994). Involved in information retrieval since the early TREC days, he has always been keen on large scale solutions to natural language processing problems, co-editing with Adam Kilgarriff a special issue of "Computational Linguistics" in 2003. Former chief scientist at the Xerox Research Centre Europe (93-01), at Clairvoyance Corporation (01-04), and with the French CEA (04-08), he has been active in transferring research into products as inventor in 16 U.S patents. In recent years, he has been working with Adrian Popescu on Geographical Indexing. With Dassault Systemes, he is interested in exploring multidimensional access to information.

Temporal Dynamics and Information Retrieval

Dr. Susan Dumais
Microsoft Research, USA
http://research.microsoft.com/en-us/um/people/sdumais/

Time: October 29, 13:30-14:30
Location: Imperial Room

Abstract

Many digital resources, like the Web, are dynamic and ever-changing collections of information. However, most of the tools information retrieval and management that have been developed for interacting with Web content, such as browsers and search engines, focus on a single static snapshot of the information. In this talk, I will present analyses of how Web content changes over time, how people re-visit Web pages over time, and how re-visitation patterns are influenced by changes in user intent and content. These results have implications for many aspects of information retrieval and management including crawling, ranking and information extraction algorithms, result presentation, and evaluation. I will describe a prototype system that supports people in understanding how the information they interact with changes over time, and a new retrieval model that incorporates features about the temporal evolution of content to improve core ranking. Finally, I will conclude with an overview of some general challenges that need to be addressed to fully incorporate temporal dynamics in information retrieval and information management systems.

Biography

Susan Dumais is a Principal Researcher and manager of the Context, Learning and User Experience for Search (CLUES) Group at Microsoft Research. Prior to joining Microsoft Research, she was at Bellcore and Bell Labs for many years, where she worked on Latent Semantic Indexing (a statistical method for concept-based retrieval), interfaces for combining search and navigation, and organizational impacts of new technology. Her current research focuses on user modeling and personalization, context and information retrieval, temporal dynamics of information, interactive retrieval, and novel evaluation methods. She has worked closely with several Microsoft groups (Bing, Windows Desktop Search, SharePoint Portal Server, and Office Online Help) on search-related innovations. Susan has published more than 200 articles in the fields of information science, human-computer interaction, and cognitive science, and holds several patents on novel retrieval algorithms and interfaces. Susan is also an adjunct professor in the Information School at the University of Washington. She is Past-Chair of ACM's Special Interest Group in Information Retrieval (SIGIR), and serves on several editorial boards, technical program committees, and government panels. She was elected to the CHI Academy in 2005, an ACM Fellow in 2006, and received the Gerard Salton Award from SIGIR for Lifetime Achievement in 2009.

Industry Track Invited Talks

The Shift from Search to Matching

Dr. Vasco Pedro
CEO and Co-founder, Bueda, Inc

Industry Session: Information Retrieval Applications (4C)
Time: October 28, 10:00-10:45
Location: British Columbia Room

Abstract

The focus on providing a personalized online experience is creating a shift in information retrieval from search strategies, where a user starts as an anonymous entity and the search engine performs a passive role, to matching strategies, where there is a wide array of information about the user even before the first interaction has taken place. In a matching driven model, the search engine takes a proactive role, functioning as a broker of information between two rich information sets. The trend can be seen in a wide range of online activities. From watching movies to online purchases, from the way we listen to music to the way we find our significant others a trend in personalization emerges as user information becomes available. As the amount of personal information online increases exponentially, the ability to understand that content is providing a wealth of opportunities for companies to provide unique online experiences in previously inaccessible areas such as search and targeted advertising. In this talk we will discuss key challenges facing information retrieval as it moves through this shift and how semantic information can help alleviate those challenges.

Biography

Vasco Pedro is the Founder and CEO of Bueda, semantic web startup focused in understanding the social web. At Bueda, he was responsible for producing the research that led to the creation of the company, building the current team and successfully guiding Bueda through two rounds of funding and two product launches. His main interests are the leveraging of semantic understanding on the social web to create smarter applications and the understanding of human behavior through semantic analysis. Vasco got a PhD in 2009 from the Language Technologies Institute, Carnegie Mellon University, with the thesis topic of Federated Ontology Search. He has authored several papers in international conferences and is currently an invited researcher at Carnegie Mellon University in addition to his role at Bueda.

Turing Dark Clouds into Silver Linings

Dr. Glenn Paulley
Director, Engineering at Sybase iAnywhere

Industry Session: Database and OLAP (5C)
Time: October 28, 13:30-14:15
Location: British Columbia Room

Abstract

Data management services are ubiquitous in the industry. With some important exceptions, relational database systems are the platform upon which many applications depend, including mainframe servers, web browsers, and handheld devices. The vast majority of these installations run unattended. Consequently, self-management, self-tuning, and self-healing features are of great importance to these systems. Cloud computing architectures, due to their inherent dynamics, add another level of flexibility - and complexity - to the problems of database self-management.

However, application developers continue to experience unpredictable performance and reliability issues with the application's software stack. Over the past few years, these issues have led to a variety of proposals to address the problem, including weak consistency models and the abandonment of SQL as a data management sub-language.

In this talk I'll present an overview of the data management problems in relational database systems that are exacerbated by cloud computing architectures, discuss the state-of-the-art in self-management technology, and conclude with some ideas for future research to address these problems.

Biography

Glenn Paulley is a Director with Sybase iAnywhere Engineering, where he manages the research and development team responsible for query processing in SQL Anywhere, Sybase's self-managing relational database server. He joined Sybase iAnywhere in 1995. During his 20-year industrial career Glenn has held previous positions at Amdahl Corporation and at a large Canadian insurance company. He holds a Ph.D in Computer Science from the University of Waterloo. His research interests include software usability, query optimization, information systems architecture, design of Management Information Systems, topics in systems analysis, interfaces to database systems, database query languages, user models, multidatabase systems, and indexing techniques.

On-line Social Networks: Insight, Commercial Value and Computational Challenges

Dr. Gordon Sun
Chief Scientist, Tencent Technology

Industry Session: Web and Social Networks (6C)
Time: October 29, 10:00-10:45
Location: British Columbia Room

Abstract

In China, Tencent is almost a ubiquitous synonym of the Internet. With over 600M active user accounts and boosting over 100M users online simultaneously, Tencent offers a unique and comprehensive window into the Chinese Internet users - the world's largest Internet population. We first present our findings studying the individual characteristics of the Tencent users and their social properties, comparing with the general Chinese Internet population. For two notable Tencent SNS products, QZone (Facebook comparable) and Weibo (Twitter comparable), we also draw comparison on various perspectives with other social networks such as the Twitter network and the coauthorship network in Computer Science. We then examine the functionality and revenue models of social networks in general. Discussed are specific details on a number of computational aspects, such as user models (behavior, interests, values, and influences), recommender systems for topics, content, and friends, social search, advertising and marketing in social networks. To conclude this talk, we list some challenging problems that we believe are calling for better techniques and solutions from the research community.

Biography

Gordon Sun, Chief Scientist, Tencent Technology, China, has been working on algorithmic search technology since 1998 when he joined Inktomi (the leading US search engine company during 1990s) as the senior scientist and architect. He also worked for two search engine companies, WiseNut and LookSmart as the director of R&D during 2001 to 2003 before he joined Yahoo! early 2004 where he was leading the Global Search Relevance team as the director of research until 2009. Gordon graduated from University of Science and Technology of China and went to the US through the CUSPEA program (sponsored by Nobel Price Winner Prof. T.D. Lee) in 1981 and received his Ph.D. in theoretical physics from University of Iowa, 1984. He worked as the Research faculty in University of Maryland for 9 years before he moved to Silicon Valley, 1993, and worked in Communication Intelligence Inc, a leading hand-writing recognition provider, as the Chief Scientist. He has broad work experiences, knowledge and publications in neural networks, pattern recognition, machine learning, data mining, information retrieval, speech recognition, hand-writing recognition and non-linear dynamics.

Digital Advertising: Personalization, Personalization, and Personalization

Dr. James Shanahan
Independent Consultant, Church and Duncan Group, Inc

Industry Session: Text Mining, Analytics, and Advertising Session (7C)
Time: October 29, 15:00-15:45
Location: British Columbia Room

Abstract

Online advertising is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers. Examples of online advertising include text ads that appear on search engine results pages, banner ads, in-text ads, or Rich Media ads that appear on regular web pages, portals or applications. Since it inception over 15 years ago, online advertising has grown rapidly and currently accounts for 13% of the overall advertising spend (which is approximately $700 billion worldwide)). A large part of the more recent success in this field has come from the following key factors:

- Personalization: offline advertising (via broadcast TV, radio, newspaper etc.) is largely a broadcast form of communication where as digital advertising is much more targeted and thus enables a personalized, and possibly informative, message to consumers.
- Interactivity: internet advertising is becoming increasingly interactive with the advent of new forms of advertising such as social advertising; this is enables advertises and consumers to operate in a more conversant manner.
- Engagement: consumers are spending more time online than with any other form of media thereby enabling a broader reach and deeper connection with consumers.
- Explainability: advertisers are beginning to understand their consumers better.

This shift in focus in digital advertising from location (i.e., publisher web pages) to personalization has brought with it numerous challenges some of which have received a lot of research attention in the DIK communities (Data, Information and Knowledge communities) over the past 10-20 years. In this talk I will review, along the dimensions outlined above, some of these key technical problems and challenges that arise when adverting becomes personal. This will be done within the context of the elaborate (and ever-evolving) ecosystems of modern day digital advertising where one has to capture, store, and process petabytes of data within the constraints of a, sometimes, sequential workflow. The ultimate goal to is provide millisecond-based decision-making at each step of this workflow that enables customizable and engaging consumer experiences.

Bibography

Jimi has spent the last 20 years developing and researching cutting-edge information management systems to harness information retrieval, linguistics and machine learning. Prior to being an independent consultant, Jimi was Chief Scientist (and member of executive team) at Turn Inc. where he focused on the development and deployment of an online ad targeting system (CPA/CPC/CPM-based) in a principled and measured way that leveraged advanced statistical and machine learning techniques. These responsibilities included leveraging the entire reservoir of data assets in order to develop

methods for identifying key optimizations, deploying relevant analytical tools and improving the user experience. Prior to joining Turn, Jimi was Principal Research Scientist at Clairvoyance Corporation where he led the Knowledge Discovery from Text Group. Before that he was a Research Scientist at Xerox Research Center Europe (XRCE), where, as a member of the Co-ordination Technologies Group, he developed Document Souls, a patented document-centric approach to information access. In the early 90s, he worked on the AI Team within the Mitsubishi Group in Tokyo.

He has published six books and over 50 research publications in the area of machine learning and information processing. Jimi was General Chair for CIKM 2008. Jimi received his Ph.D. in engineering mathematics from the University of Bristol, U. K. and holds a bachelor of science degree in computer science from the University of Limerick, Ireland. He is a Marie Curie fellow and member of IEEE and ACM.

CIKM'10 Main Conference Program

Wednesday, October 27, 2010

Conference Opening (October 27, 8:10-8:30)
Room: Imperial
Welcome: Conference Chair
Program Introduction: PC Chairs

Keynote Talk (October 27, 8:30-9:30)
Chair: Jimmy Huang
Room: Imperial
Title: Search Engine Support for Software Applications
Speaker: Jamie Callan, *Carnegie Mellon University*

Coffee Break (9:30-10:00, Imperial Foyer)

Session 1A (October 27, 10:00-12:00)
KM Track: Information Extraction
Chair: Charles Clarke
Room: Alberta

- Components for Information Extraction: Ontology-Based Information Extractors and Generic Platforms, *Daya Wimalasuriya and Dejing Dou*
- Automatically Acquiring a Semantic Network of Related Concepts, *Sean Szumlansku and Fernando Gomez*
- Online Annotation of Text Streams with Structured Entities, *Ken Pu, Richard Drake, Oktie Hassanzadeh and Renee Miller*
- Automatic Extraction of Web Data Records Containing User-Generated Context, *Xinying Song, Jing Liu, Yunbo Cao and Chin-Yew Lin*
- An Effective Algorithm for Mining Time Interval-based Patterns in Large Database, *Yi-cheng, Ji-Chiang Jiang, Wen-Chih Peng and Suh-Yin Lee*

Session 1B (October 27, 10:00-12:00)
IR Track: Ranking and Retrieval Model
Chair: Jian-Yun Nie
Room: Quebec

- What can Quantum Theory Bring to Information Retrieval? *Benjamin Piwowarski, Ingo Frommholz, Mounia Lalmas and Keith van Rijsbergen*
- Entity Ranking using Wikipedia as a Pivot, *Rianne Kaptein, Jaap Kamps, Pavel Serdyukov and Arjen de Vries*
- Ranking under Temporal Constraints, *Lidan Wang, Donald Metzler and Jimmy Lin*
- Examining the Information Retrieval Process from an Inductive Perspective, *Ronan Cummins, Mounia Lalmas and Colm O'Riordan*
- On Identifying Representative Relevant Documents, *Fiana Raiber and Oren Kurland*

Session 1C (October 27, 10:00-12:00)
DB Track: Indexes and Query Optimization
Chair: David Cheung
Room: British Columbia

- On the Selectivity of Multidimensional Routing Indices, *Christos Doulkeridis, Akrivi Vlachou, Kjetil Noervaag, Yannis Kotidis and Michalis Vazirgiannis*
- Path-Hop: Efficiently Indexing Large Graphs for Reachability Queries, *Jing Cai and Chung Keung Poon*
- On Wavelet Decomposition of Uncertain Time Series Data Sets, *Yuchen Zhao, Charu Aggarwal and Philip Yu*
- Efficient Set-Correlation Operator inside Databases, *Shaoxu Song and Lei Chen*
- Online Update of B-trees, *Marina Barsky, Alex Thomo, Zoltan Toth and Calisto Zuzarte*

Session 1D (October 27, 10:00-12:00)
IR Track: Domain-Specific and Multimedia IR
Chair: Gareth Jones
Room: Tudor 7 & 8

- Wisdom of the Ages: Toward Delivering the Children's Web with the Link-based AgeRank Algorithm, *Karl Gyllstrom and Marie-Francine Moens*
- A Cross-lingual Framework for Monolingual Biomedical Information Retrieval, *Dolf Trieschnigg, Djoerd Hiemstra, Franciska de Jong and Wessel Kraaij*
- Multi-Modal Multi-Correlation Personal News Retrieval, *Zechao Li, Jing Liu, Xiaobin Zhu and Hanqing Lu*
- Bringing Order to Your Photos: Event-Driven Classification of Flickr Images Based on Social Knowledge, *Claudiu Firan, Mihai Georgescu, Wolfgang Nejdl and Raluca Paiu*

Lunch and Posters (12:00-13:30)
Room: Imperial

Session 2A (October 27, 13:30-15:30)
KM Track: Link and Graph Mining
Chair: Benjamin Fung
Room: Alberta

- Mining Topic-Level Influence in Heterogeneous Networks, *Lu Liu, Jie Tang, Meng Jiang, Jiawei Han and Shiqiang Yang*
- Mining Interesting Link Formation Rules in Social Networks, *Cane Wing-ki Leung, Ee-Peng Lim, David Lo and Jianshu Weng*
- SHRINK: A Structural Clustering Algorithm for Detecting Hierarchical Communities in Networks, *Jianbin Huang, Heli Sun, Jiawei Han, Hongbo Deng and Yizhou Sun*
- Outcome Aware Ranking in Interaction Networks, *S Kameshwaran, Sameep Mehta and Vinayaka Pandit*
- Expansion and Search in Networks, *Arun Maiya and Tanya Berger-Wolf*

Session 2B (October 27, 13:30-15:30)
IR Track: Machine Learning for IR (I)
Chair: David Grossman
Room: Quebec

- Improved Latent Concept Expansion Using Hierarchical Markov Random Fields, *Hao Lang, Donald Metzler, Bin Wang and Jin-Tao Li*
- Term Necessity Prediction, *Le Zhao and Jamie Callan*
- Decomposing Background Topics from Keywords by Principal Component Pursuit, *Kerui Min, Zhengdong Zhang, John Wright and Yi Ma*
- Document Update Summarization using Incremental Hierarchical Clustering, *Dingding Wang and Tao Li*
- How about utilizing ordinal information from the distribution of unlabeled data? *Mingjie Qian, Bo Chen and Hongwei Qi*

Session 2C (October 27, 13:30-15:30)
DB Track: Mobile and Distributed Data Management
Chair: Frank Tompa
Room: British Columbia

- Automatic Schema Merging Using Mapping Constraints over Incomplete Sources, *Xiang Li, Christoph Quix, David Kensche and Sandra Geisler*
- Preserving Location and Absence Privacy in Geo-Social Networks, *Dario Freni, Carmen Ruiz Vicente, Sergio Mascetti, Claudio Bettini and Christian S. Jensen*
- Preference Query Evaluation Over Expensive Attributes, *Justin Levandoski, Mohamed Mokbel and Mohamed Khalefa*
- Energy-Efficient Top-k Query Processing in Wireless Sensor Networks, *Baichen Chen, Weifa Liang, Rui Zhou and Jeffrey Yu*
- StableBuffer: Optimizing Write Performance for DBMS Applications on Flash Devices, *LI YU, Xu Jianliang, Byron CHOI and Haibo Hu*

Session 2D (October 27, 13:30-15:30)
KM Track: Classification and Clustering
Chair: Dominic Slezak
Room: Tudor 7 & 8

- Mr.KNN: Soft Relevance for Multi-label Classification, *Xiaotong Lin and Xue-wen Chen*
- Collaborative Dual-PLSA: Mining Distinction and Commonality across Multiple Domains for Classification, *Fuzhen Zhuang, Ping Luo, Zhiyong Shen, Qing He, Yuhong Xiong, Zhongzhi Shi and Hui Xiong*
- Inferring Gender of Movie Reviewers: Exploiting Writing Style, Content and Metadata, *Jahna Otterbacher*
- A Robust Semi-supervised Classification Method for Transfer Learning, *Akinori Fujino, Naonori Ueda and Masaaki Nagata*
- Multi-View Clustering with Constraint Propagation for Learning with an Incomplete Mapping Between Views, *Eric Eaton, Marie desJardins and Sara Jacob*

Coffee Break (15:30-16:00, British Columbia Foyer)

Session 3A (October 27, 16:00-17:40)
KM Track: Large-scale Statistical Techniques
Chair: Chris Clifton
Room: Alberta

- Pricing Guaranteed Contracts in Online Display Advertising, *Vijay Bharadwaj, Wenjing Ma, Michael Schwarz, Jayavel Shanmugasundaram, Erik Vee, Jack Xie and Jian Yang*
- Maximum Normalized Spacing for Efficient Visual Clustering, *Zhi-Gang Fan, Yadong Wu and Bo Wu*
- Multilevel Manifold Learning with Application to Spectral Clustering, *Haw-ren Fang, Sophia Sakellaridi and Yousef Saad*
- Accelerating Probabilistic Frequent Itemset Mining: A Model-Based Approach, Liang Wang , *Reynold Cheng, Sau Dan Lee and David Cheung*

Session 3B (October 27, 16:00-17:40)
IR Track: Machine Learning for IR (II)
Chair: Irwin King
Room: Quebec

- Learning Click Models via Probit Bayesian Inference, *Yuchen Zhang, Gang Wang and Weizhu Chen*
- Document Allocation Policies for Selective Searching of Distributed Indexes, *Anagha Kulkarni and Jamie Callan*
- Rank Learning for Factoid Question Answering with Linguistic and Semantic Constraints, *Matthew Bilotti, Jonathan Elsas, Jaime Carbonell and Eric Nyberg*
- Learning to Rank Relevant and Novel Documents through User Feedback, *Abhimanyu Lad and Yiming Yang*

Session 3C (October 27, 16:00-17:40)
DB Track: Top-K and Shortest Path Processing
Chair: Xiaohui Yu
Room: British Columbia

- Set Cover Algorithms For Very Large Datasets, *Graham Cormode, Howard Karloff and Anthony Wirth*
- The Gist of Everything New: Personalized Top-k Processing over Web 2.0 Streams, *Parisa Haghani, Sebastian Michel and Karl Aberer*
- Fast and Accurate Estimation of Shortest Paths in Large Graphs, *Andrey Gubichev, Srikanta Bedathur, Stephan Seufert and Gerhard Weikum*
- Fast Top-k Simple Shortest Paths Discovery on Graphs, *Gao Jun , Huida Qiu and Xiao Jiang*

Session 3D (October 27, 16:00-17:40)
IR Track: IR Evaluation
Chair: Susan Dumais
Room: Tudor 7 & 8

- Factors Affecting Click-Through Behavior in Aggregated Search Interfaces, *Shanu Sushmita, Hideo Joho, Mounia Lalmas and Robert Villa*
- Web Search Solved? All Result Rankings the Same? *Hugo Zaragoza, B. Barla Cambazoglu and Ricardo Baeza-Yates*
- Assessor Error in Stratified Evaluation, *William Webber, Doug Oard, Falk Scholer and Bruce Hedin*
- CiteData: A New Multi-faceted Dataset for Evaluating Personalized Search Performance, *Abhay Harpale, Yiming Yang, Siddharth Gopal, Daqing He and Zhen Yue*

Conference Reception and Posters (October 27, 18:00-21:00)
Room: Imperial

Thursday, October 28, 2010

Keynote Talk (October 28, 8:30-9:30)
Chair: Nick Koudas
Room: Imperial
Title: Schema Extraction
Speaker: Divesh Srivastava, *AT&T Labs-Research*

Coffee Break (9:30-10:00, Imperial Foyer)

Session 4A (October 28, 10:00-12:00)
KM Track: Information Filtering and Recommender Systems (I)
Chair: Arjen P. De Vries
Room: Alberta

- Learning a User-Thread Alignment Manifold for Thread Recommendation in Online Forum, *Jun Zhao, Jiajun Bu, Chun Chen, Ziyu Guan and Can Wang*
- FacetCube: A Framework of Incorporating Prior Knowledge into Non-negative Tensor Factorization, *Yun Chi and Shenghuo Zhu*
- Travel Route Recommendation using Geotags in Photo Sharing Sites, *Takeshi Kurashima, Tomoharu Iwata, Go Irie and Ko Fujimura*
- Boosting Social Network Connectivity with Link Revival, *Yuan Tian, Qi He, Qiankun Zhao, Xingjie Liu and Wang-Chien Lee*
- Power in Unity: Forming Teams in Large-Scale Community Systems, *Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis and Stefano Leonardi*

Session 4B (October 28, 10:00-12:00)
IR Track: Social Networks and Text Mining
Chair: Yi Chang
Room: Quebec

- Who Should I Cite? Learning Literature Search Models from Citation Behavior, *Steven Bethard and Dan Jurafsky*
- A Structured Approach to Query Recommendation With Social Annotation Data, *Jiafeng Guo, Xueqi Cheng, Gu Xu and Huawei Shen*
- Using the Past to Score the Present: Extending Term Weighting Models with Revision History Analysis, *Ablimit Aji, Yu Wang, Eugene Agichtein and Evgeniy Gabrilovich*
- Bag-of-Word Pyramid and Multi-Scale Text Analysis, *Shuang-Hong Yang and Hongyuan Zha*
- Latent Interest-Topic Model: Finding the causal relationships behind dyadic data, *Noriaki Kawamae*

Session 4C (October 28, 10:00-12:00)
Industry Track: IR Applications
Chair: Paul Bennett
Room: British Columbia

- **Invited Talk:** The shift from search to matching, *Vasco Pedro, CEO and Co-founder, Bueda, Inc.*
- **PROSPECT**: A system for screening candidates for recruitment, *Amit Singh, Rose Catherine, Karthik Visweswariah, Vijil Chenthamarakshan and Nanda Kambhatla*
- Active Caching for Similarity Queries Based on Shared-Neighbor Information, *Michael Houle, Vincent Oria and Umar Qasim*

- Searching Consumer Image Collections Using Web-based Concept Expansion, *Mark Wood, Stacie Hibino and Alexander Loui*

Session 4D (October 28, 10:00-12:00)
DB Track: Information Retrieval in Databases
Chair: Leif Azzopardi
Room: Tudor 7 & 8

- Index Structures for Efficiently Searching Natural Language Text, *Pirooz Chubak and Davood Rafiei*
- Efficient Temporal Keyword Queries over Versioned Text, *Avishek Anand, Srikanta Bedathur, Klaus Berberich and Ralf Schenkel*
- Result-Size Estimation for Information-Retrieval Subqueries, *Guido Sautter, Klemens Böhm and Andranik Khachatryan*
- FACeTOR: Cost-Driven Exploration of Faceted Query Results, *Abhijith Kashyap, Vagelis Hristidis and Michalis Petropoulos*
- A Framework for Evaluating Database Keyword Search Strategies, *Joel Coffman and Alfred Weaver*

Lunch (12:00-13:30, on your own)
[Participants will need to provide their own lunches]

Session 5A (October 28, 13:30-15:30)
KM Track: Temporal, Spatial and Stream Data Mining
Chair: Parke Godfrey
Room: Alberta

- Network Growth and the Spectral Evolution Model, Jérôme Kunegis, *Damien Fay and Christian Bauckhage*
- Robust Automatic Detection of Craters in Planetary Images, *Wei Ding, Tomasz Stepinski, Lourenco Bandeira, Ricardo Vilalta, Youxi Wu, Zhenyu Lu and Tianyu Cao*
- You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users, *Zhiyuan Cheng, James Caverlee and Kyumin Lee*
- Partial Drift Detection Using a Rule Induction Framework, *Damon Sotoudeh and Aijun An*
- A Method for Discovering Components of Human Rituals from Streams of Sensor Data, *Bamis Athanasios, Jia Fang and Andreas Savvides*

Session 5B (October 28, 13:30-15:30)
IR Track: Filtering and Recommendation
Chair: Ingmar Weber
Room: Quebec

- Two-Tier Similarity Model for Story Link Detection, *Tadashi Nomoto*
- Selected new training documents to update user profile, *Abdulmohsen Algarni, Yuefeng Li and Yue Xu*
- Collaborative Filtering in Social Tagging Systems Based on Joint Item-Tag Recommendations, *Jing Peng, Daniel Zeng, Huimin Zhao and Fei-Yue Wang*
- Collaborative Future Event Recommendation, *einat minkov, Ben Charrow, Jonathan Ledlie, Seth Teller and Tommi Jaakkola*
- Hybrid Tag Recommendation for Social Annotation Systems, *Jonathan Gemmell, Thomas Schimoler, Bamshad Mobasher and Robin Burke*

Session 5C (October 28, 13:30-15:30)
Industry Track: Databases and OLAP
Chair: Robert Warren
Room: British Columbia

- **Invited Talk:** Turning Dark Clouds into Silver Linings, *Glenn Paulley, Director, Engineering, Sybase iAnywhere*
- XML Schema Computations: Schema Compatibility Testing and Subschema Extractions, *Thomas Lee and David Cheung*
- Visual Cube and On-Line Analytical Processing of Images, *Xin Jin, Liangliang Cao, Jiebo Luo, Bolin Ding, Xide Lin and Jiawei Han*
- Pattern Discovery for Large Mixed-Mode Databases, *Andrew K.C. Wong, Bin Wu, Gene P.K. Wu and Chan Keith*

Session 5D (October 28, 13:30-15:30)
KM Track: Data Pre- and Post- Processing
Chair: Xue-Wen Chen
Room: Tudor 7 & 8

- Constructing Classification Features using Minimal Predictive Patterns, *Iyad Batal and Milos Hauskrecht*
- RankSVR: Can Preference Data Help Regression? *Hwanjo Yu, Sungchul Kim and Seung-Hoon Na*
- Estimating Accuracy for Text Classification Tasks on Large Unlabeled Data, *Snigdha Chaturvedi, Tanveer Faruquie, Venkata Subramaniam and Mukesh Mohania*
- A Probabilistic Topic-Connection Model for Automatic Image Annotation, *Xin Chen, Xiaohua Hu, Zhongna Zhou, Caimei Lu, Gail Rosen, Tingting He and E.K. Park*
- Orientation Distance-based Discriminative Feature Extraction for Multi-Class Classification, *Bo Liu, Yanshan Xiao and Longbing Cao*

Coffee Break (15:30-16:00, Imperial Foyer)

Demo Session A: IR [I] (October 28, 16:00-17:30)
Chair: Cathal Gurrin
Room: Imperial

- EntityEngine: Answering Entity-Relationship Queries using Shallow Semantics, *Xiaonan Li, Chengkai Li and Cong Yu*
- Facetedpedia: Enabling Query-Dependent Faceted Search for Wikipedia, *Ning Yan, Chengkai Li, Senjuti B. Roy, Rakesh Ramegowda and Gautam Das*
- Discovering, Ranking and Annotating Cross-Document Relationships between Concepts, *Wei Jin and Xin Wu*
- TC-DCA: A System for Text Classification Based on Document's Content Allocation, *Wenbo Li, Le Sun, Zhenzhong Zhang, Xue Jiang and Weiru Zhang*
- XReal: An Interactive XML Keyword Searching, *Zhifeng Bao, Jiaheng Lu and Tok Wang Ling*

Demo Session B: IR [II] (October 28, 16:00-17:30)
Chair: Cathal Gurrin
Room: Imperial

- EUI: An Embedded Engine for Understanding User Intents from Mobile Devices, *SangKeun Lee, JongWoo Ha, Jung-Hyun Lee, Kyu-Sun Shim and Namhoon Kim*

- WikiPop - Personalized Event Detection System Based on Wikipedia Page View Statistics, *Marek Ciglan and Kjetil Nørvåg*
- Summarizing biological literature with BioSumm, *Elena Baralis and Alessandro Fiori*
- Crawling the Web for Structured Documents, *Julián Urbano, Juan Llorens, Yorgos Andreadakis and Monica Marrero*
- Connecting the local and the online in information management, *Gabriella Kazai, Natasa Milic-Frayling, Tim Haughton, Natalia Manola, Katerina Iatropoulou, Antonis Lempesis, Paolo Manghi and Marko Mikulicic*

Demo Session C: DB (October 28, 16:00-17:30)
Chiar: Cathal Gurrin
Room: Imperial

- LiquidXML: Adaptive XML content redistribution, *Jesus Camacho-Rodriguez, Asterios Katsifodimos, Alexandra Roatis and Ioana Manolescu*
- Brown Dwarf: A P2P Data-Warehousing System, *Katerina Doka, Dimitrios Tsoumakos and Nectarios Koziris*
- RDFViewS: A Storage Tuning Wizard for RDF Applications, *Francois Goasdoue, Konstantinos Karanasos, Julien Leblay and Ioana Manolescu*
- SEQUEL: Query Completion via Pattern Mining on Multi-Column Structural Data, *Chuancong Gao, Qingyan Yang and Jianyong Wang*
- MI-WDIS: Web Data Integration System for Market Intelligence, *Zhongmin Yan, Qingzhong Li, Zhaohui Peng, Yongquan Dong, Yanhui Ding, Yongxin Zhang and Xiuxing Xu*

Demo Session D: KM (October 28, 16:00-17:30)
Chair: Cathal Gurrin
Room: Imperial

- FALCON: Seamless Access to Meeting Data from the Inbox and Calendar, *Jennifer Lai, Jie Lu, Mercan Topkara, Mukundan Desikan, Yi Guo, Ajinkya Kale, Nizar Lethif, Stephan Wissel, Peter Bjellerup and Karl Cama*
- WS-GraphMatching: A Web Service Tool for Graph Matching, *Qiong Cheng, Jinpeng Wei, Alexander Zelikovsky and Mitsunori Ogihara*
- SPAC: A Distributed; Peer-to-Peer, Secure and Privacy-aware Social Space, *Angela Bonifati, Hui Wang and Ruilin Liu*
- i-SEE: Integrated Stream Execution Environment over On-line Data Streams, *Se Jung Shin, Hong Kyu Park, Ho Jin Woo and Won Suk Lee*
- Exploring and Visualizing Academic Social Networks, *Veselin Ganev, Zhaochen Guo, Diego Serrano, Denilson Barbosa and Eleni Stroulia*

Conference Banquet (October 28, 18:30-21:30)
Location: Governor's Room, Liberty Grand, 25 British Columbia Road, Exhibition Place, Toronto.
Transportation: Buses will leave Royal York Hotel at 17:45 and 18:15.

Friday, October 29, 2010

Keynote Talk (October 29, 8:30-9:30)
Chair: Gareth Jones
Room: Imperial
Title: Use of Semantics in Real Life Applications
Speaker: Gregory Grefenstette, Exalead

Coffee Break (9:30-10:00, Imperial Foyer)

Session 6A (October 29, 10:00-12:00)
KM Track: Information Filtering and Recommender Systems (II)
Chair: Jie Tang
Room: Alberta

- Evaluating, Combining and Generalizing Recommendations with Prerequisites, *Aditya Parameswaran, Hector Garcia-Molina and Jeffrey Ullman*
- Automatically Suggesting Topics for Augmenting Text Documents, *Robert West, Doina Precup and Joelle Pineau*
- Detecting Product Review Spammers using Rating Behaviors, *Ee-Peng Lim, Bing Liu, Viet-An Nguyen, Nitin Jindal and Hady Lauw*
- Classical Music for Rock Fans?: Novel Recommendations for Expanding User Interests, *Nakatsuji Makoto, Yasuhiro Fujiwara, Akimichi Tanaka, Toshio Uchiyama, Ko Fujimura and Toru Ishida*
- Improving One-Class Collaborative Filtering by Incorporating Rich User Information, *Yanen Li, Jia Hu, Chengxiang Zhai and Ye Chen*

Session 6B (October 29, 10:00-12:00)
IR Track: User Modeling and Search Personalization
Chair: Noriko Kando
Room: Quebec

- Personalized Search by Tag-based User Profile and Resource Profile in Collaborative Tagging Systems, *Yi Cai and Qing Li*
- A Comparison of User and System Query Performance Predictions, *Claudia Hauff, Diane Kelly and Leif Azzopardi*
- The Anatomy of a Click: Modeling user behavior on web information systems, *Kunal Punera*
- Exploring Online Social Activities for Adaptive Search Personalization, *Qihua Wang and Hongxia Jin*
- Predicting Short-Term Interests Using Activity-Based Search Context, *Ryen White, Paul Bennett and Susan Dumais*

Session 6C (October 29, 10:00-12:00)
Industry Track: Web and Social Networks
Chair: Kevyn Collins-Thompson
Room: British Columbia

- **Invited Talk:** On-line Social Networks: Insight, Commercial Value and Computational Challenges, *Gordon Sun, Chief Scientist, Tencent Technology*
- Probabilistic First Pass Retrieval for Search Advertising: From Theory to Practice, *Hema Raghavan and Rukmini Iyer*

- Faceted Search and Browsing of Audio Content on Spoken Web, *Mamadou Diao, Sougata Mukherjea, Nitendra Rajput and Kundan Srivastava*
- Predicting Product Adoption in Large-Scale Social Networks, *Rushi Bhatt, Vineet Chaoji and Rajesh Parekh*

Session 6D (October 29, 10:00-12:00)
IR Track: Query Analysis and Feedback
Chair: Elizabeth Liddy
Room: Tudor 7 & 8

- Reverted Indexing for Feedback and Expansion, *Jeremy Pickens, Matthew Cooper and Gene Golovchinsky*
- Improving Verbose Queries using Subset Distribution, *Xiaobing Xue, Samuel Huston and Bruce Croft*
- A unified optimization framework for robust, risk-aware pseudo-relevance feedback, *Joshua Dillon and Kevyn Collins-Thompson*
- Ranking Related Entities: Components and Analyses, *Marc Bron, Krisztian Balog and Maarten de Rijke*

Lunch (12:00-13:30)
Location: Ballroom, 3rd floor

Keynote Talk (October 29, 13:30-14:30)
Chair: Kevyn Collins-Thompson
Room: Imperial
Title: Temporal Dynamics and Information Retrieval
Speaker: Susan Dumais, *Microsoft Research*

Coffee Break (14:30-15:00, Imperial Foyer)

Session 7A (October 29, 15:00-17:00)
KM Track: Semantic Techniques
Chair: Aijun An
Room: Alberta

- Semantic tags generation and retrieval for online advertising, *Mirizzi Roberto, Azzurra Ragone, Tommaso Di Noia and Eugenio Di Sciascio*
- MENTA: Inducing Multilingual Taxonomies from Wikipedia, *Gerard de Melo and Gerhard Weikum*
- Ontology Emergence from Folksonomies, *Kaipeng Liu, Binxing Fang and Weizhe Zhang*
- Multi-document Topic Segmentation, *Minwoo Jeong and Ivan Titov*
- Meta-Metadata: A Metadata Semantics Language for Collection Representation Applications, *Andruid Kerne, Yin Qu, Andrew Webb, Sashikanth Damaraju, Nic Lupfer and Abhinav Mathur*

Session 7B (October 29, 15:00-17:00)
IR Track: Web Search
Chair: Jaap Kamps
Room: Quebec

- Clickthrough-Based Translation Models for Web Search: from Word Models to Phrase Models, *Jianfeng Gao, Xiaodong He and Jian-Yun Nie*

- Temporal Query Log Profiling to Improve Web Search Ranking, *Alexander Kotov, Pranam Kolari, Lei Duan and Yi Chang*
- Improving Web Search Relevance and Freshness with Content Previews, *Siva Gurumurthy, Hang Su, Vasileios Kandylas and Vidhyashankar Venkataraman*
- Organizing Query Completions for Web Search, *Alpa Jain and Gilad Mishne*
- Selectively Diversifying Web Search Results, *Rodrygo Santos, Craig Macdonald and Iadh Ounis*

Session 7C (October 29, 15:00-17:00)
Industry Track: Text Mining and Analytics
Chair: Gregory Grefenstette
Room: British Columbia

- **Invited Talk:** Digital Advertising: Personalization, Personalization, and Personalization, *James Shanahan, Independant Consultant, Church and Duncan Group, Inc.*
- Building Re-usable Dictionary Repositories for Real-world Text Mining, *Shantanu Godbole, Indrajit Bhattacharya, Ajay Gupta and Ashish Verma*
- OpinionIt: A Text Mining System for Cross-lingual Opinion Analysis, *Honglei Guo and Huijia Zhu*
- Hierarchical Service Analytics for Improving Productivity in an Enterprise Service Center, *Chunye Wang, Ram Akella and Srikant Ramachandran*

Session 7D (October 29, 15:00-17:00)
IR Track: Scalability and Efficiency
Chair: Frank Tompa
Room: Tudor 7 & 8

- VSEncoding: Efficient Coding and Fast Decoding of Integer Lists via Dynamic Programming, *Rossano Venturini and Fabrizio Silvestri*
- Efficient Term Proximity Search with Term-Pair Indexes, *Hao Yan, Shuming Shi, Fan Zhang, torsten suel and Ji-Rong Wen*
- Improved Index Compression Techniques for Versioned Document Collections, *Jinru He, Junyuan Zeng, Torsten Suel*
- Building Efficient Multi-Threaded Search Nodes, *Carolina Bonacic, Carlos Garcia, Marin Mauricio, Manuel Prieto-Matias and Francisco Tirado*
- Real Time Memory Efficient Data Redundancy Removal Algorithm, *Ankur Narang, Souvik Bhattacherjee and Vikas Garg*

CIKM'10 Posters

October 27, 12:00 – 21:00 ***(Imperial Room)***

DB Track

DB1	Search-Log Anonymization and Advertisement: Are They Mutually Exclusive?	Thorben Burghardt, Klemens Bohm, Achim Guttmann, Christopher Clifton
DB2	Automated Interaction in Social Networks with Datalog	Royi Ronen, Oded Shmueli
DB3	A Data Lineage Model for Distributed Sub-image Processing	Mwebaze Johnson, John McFarland, Danny Boxhoorn, Edwin Valentine
DB4	Open User Schema Guided Evaluation of Streaming RDF Queries	Zhihong Chong, Qi Guilin, Weiwei Ni, Hu Shu, Zhou Aoying
DB5	SUMMA: Subgraph Matching in Massive Graphs	Shijie Zhang, Shirong Li, Jiong Yang
DB6	Automatically Weighting Tags in XML Collection	Dexi Liu, Changxuan Wan, Lei Chen, Xiping Liu
DB7	Skyline Query Processing for Uncertain Data	Mohamed Khalefa, Mohamed Mokbel, Justin Levandoski
DB8	FD-Buffer: A Buffer Manager for Databases on Flash Disks	Bingsheng He, sai tung on, Yinan Li, Ming Wu, Qiong Luo, Xu Jianliang
DB9	Efficiently Querying Archived Data using Hadoop	rajeev gupta, Himanshu Gupta, Ullas Nambiar, Mukesh Mohania
DB10	Evaluation of Top-k queries in Peer-to-Peer Networks using Threshold Algorithms	Ioannis Chrysakis, Constantinos Chalkidis, Dimitris Plexousakis
DB11	A Metamodel Approach to Flexible Semantic Web Service discovery	Roberto De Virgilio, Devis Bianchini
DB12	On Efficient Top-k Social Web Search	Peifeng Yin, Ken C.K. Lee, Wang-Chien Lee
DB13	Quantifying Uncertainty in Multi-Dimensional Cardinality Estimations	Andranik Khachatryan, Klemens Boehm
DB14	An AML-based Method for Entity Matching by Exploiting SoftFDs Synonyms on the Web	Zhixu Li, Laurianne Sitbon, Liwei Wang, Xiaofang Zhou, Xiaoyong Du
DB15	An Efficient Data-Centric Storage Scheme Considering Storage and Query Hot-Spots in Sensor Networks	Yonghun Park, Dongmin Seo, Jonghyeon Yun, Christopher T. Ryu, Jaesoo Yoo
DB16	Formal Approach and Automated Tool for Constructing Ontology from Object-oriented Database Model	Fu Zhang, Z. M. Ma, Gaofeng Fan Fan, Yu Wang

DB17	(k, P)-anonymity: Towards Pattern-Preserving Anonymity of Time-Series Data	Xuan Shang, Ke Chen, Lidan Shou, Gang Chen, Tianlei Hu
DB18	Concurrent Atomic Protocols for Making and Changing Decisions in Social Networks	Royi Ronen, Oded Shmueli
DB19	Extending Dictionary-based Entity Recognition to Tolerate Errors	Guoliang Li, Jianhua Feng
DB20	Yet Another Write-Optimized DBMS Layer for Flash-Based Solid State Storage	Hongchan Roh, Daewook Lee, Sanghyun Park
DB21	PrInt: A Provenance Model to Support Integration Processes	Bruno Tomazela, Carmem Satie Hara, Ricardo Rodrigues Ciferri, Cristina Ciferri
DB22	OLAP-based Query Recommendation	Carlos Garcia-Alvarado, Zhibo Chen, Carlos Ordonez
DB23	Selective Data Acquisition for Probabilistic K-NN Query	Yu-Chieh Lin, De-Nian Yang, Ming-Syan Chen
DB24	Support Elements in Graph Structured Schema Reintegration	Xun Sun, Pottinger Rachel, Michael Lawrence
DB25	BP-tree: An Efficient Index for Similarity Search in High-Dimensional Metric Spaces	Jurandy Almeida, Ricardo da Silva Torres, Neucimar Leite
DB26	Query Optimization for Ontology-Based Information Integration	Yingjie Li, Jeff Heflin
DB27	Aspect-Oriented Relations and Queries	Curtis Dyreson
DB28	A Hierarchical Approach to Reachability Query Answering in Very Large Graph Databases	Saikat Dey, Hasan Jamil
DB29	Computing the Top-k Maximal Answers in a Join of Ranked Lists	Yaron Kanza, Mirit Shalem
DB30	PruSM: A prudent Schema Matching strategy for Web Forms	Thanh Nguyen
DB31	Anonymizing Data with Quasi-Sensitive Attribute Values	Li Xiong

IR Track

IR1	Exploiting Site-Level Information to Improve Web Search	Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski, George Mavromatis, Donald Metzler, Jane Wang
IR2	Exploration-Exploitation Tradeoff in Interactive Relevance Feedback	Maryam Karimzadehgan, ChengXiang Zhai
IR3	Ranking Social Bookmarks Using Topic Models	Morgan Harvey, Mark Carman, Ian Ruthven
IR4	TABEX: Web-Scale Semi-Structured Table Census and Classification	Eric Crestan, Patrick Pantel
IR5	Challenges in Personalized Authority Flow Based Ranking of Social Media	Hassan Sayyadi, John Edmond, Vagelis Hristidis, Louiqa Raschid
IR6	A new mathematics retrieval system	Shahab Kamali, Frank Tompa
IR7	Explore Click Models for Search Ranking	Dong Wang, Weizhu Chen
IR8	Generating Advertising Keywords from Video Content	Michael Welch, Junghoo (John) Cho, Walter Chang
IR9	Finding web pages for children	Carsten Eickhoff, Pavel Serdyukov, Arjen de Vries
IR10	Rough Sets Based Reasoning and Pattern Mining for a Two-Stage Information Filtering System	Xujuan Zhou, Yuefeng Li, Peter Bruza, Yue Xu, Raymond Y.K. Lau
IR11	A Late Fusion Approach to Monolingual & Cross-lingual Document Re-ranking	Dong Zhou, Seamus Lawless, Jinming Min, Vincent Wade
IR12	Learning to Generate Summary as Structured Output	Hiroya Takamura, Manabu Okumura
IR13	Group Ranking with Application to Image Retrieval	Ou Wu
IR14	Co-Ranking: Unifying Explicit and Implicit Feedback for Collaborative Filtering	Nathan Liu
IR15	Directly Optimizing Evaluation Measures in Learning to Rank based on the Clonal Selection Algorithm	Qiang He, Jun Ma, Shuaiqiang Wang
IR16	Query Model Refinement Using Word Graphs	Yunping Huang, Le Sun, Jian-Yun Nie

IR17	Building Recommendation Systems using Peer-to-Peer Shared Content	Yuval Shavitt, Ela Weinsberg, Udi Weinsberg
IR18	Threshold Behavior of Incentives in Social Networks	Nagaraj Kota, Narahari Y.
IR19	Image Retrieval at Memory担 Edge: Known Image Search based on User-Drawn Sketches	Michael Springmann, Ihab Al Kabary, Heiko Schuldt
IR20	Utilizing Refinding for Personalized Information Retrieval	Sarah Tyler, Jian Want, Yi Zhang
IR21	User Behavior Driven Ranking without Editorial Judgments	Shihao Ji, Taesup Moon, Geroges Dupret, Ciya Liao, Zhaohui Zheng
IR22	Query Reformulation using Conditional Random Fields	Jitendra Ajmera, Hema Swetha Koppula, Krishna P. Leela
IR23	A Feature-Word-Topic model for image annotation	Cam-Tu Nguyen, Natsuda Kaothanthong, Xuan-Hieu Phan, Takeshi Tokuyama
IR24	Weighting Common Syntactic Structures for Natural Language based Information Retrieval	Liu Chang, Chang Liu, Epaminondas Kapetanios
IR25	Ranking with Auxiliary Data	Bo Long, Yi Chang, Zhaohui Zheng, Shuang-Hong Yang
IR26	Using Term Dependency According to Its Utility	Lixin Shi, Jian-Yun Nie
IR27	Modeling Reformulation Using Query Distributions	Xiaobing Xue, Bruce Croft, David Smith
IR28	Online learning for recency search ranking using real-time user feedback	Taesup Moon, Lihong Li, Wei Chu, Ciya Liao, Zhaohui Zheng, Yi Chang
IR29	Expert Identification in Community Question Answering: Exploring Question Selection Bias	Aditya Pal, Joseph Konstan
IR30	On the Relationship Between Novelty and Popularity of User-Generated Content	David Carmel, Haggai Roitman, Elad Yom-Tov
IR31	Focused Crawling Using Navigational Rank	Shicong Feng, Li Zhang, Yuhong Xiong and Conglei Yao
IR32	TAER: Time-Aware Entity Retrieval - Exploiting the Past to find Relevant Entities in News Articles	Gianluca Demartini, Malik Muhammad Saad Missen, Roi Blanco, Hugo Zaragoza

IR33	Demographich Information Flows - How web queries move between user segments	Ingmar Weber, Alejandro Jaimes
IR34	Detecting Periodic Changes in Search Intentions in a Search Engine	Masaya Murata, Hiroyuki Toda, Yumiko Matsuura, Ryoji Kataoka, Takayoshi Mochizuki
IR35	Recommendation Based On Object Typicality	Yi Cai, Ho-fung Leung, Qing Li, Jie Tang, Juanzi Li
IR36	Selecting keywords for content based recommendation	Christian Wartena, Wout Slakhorst, Martin Wibbels
IR37	Structural Annotation of Search Queries Using Pseudo-Relevance Feedback	Michael Bendersky, Bruce Croft, David Smith
IR38	Search as if You were in Your Home Town: Geographic Search by Regional Context and Dynamic Feature	Makoto Kato, Hiroaki Ohshima, Satoshi Oyama, Katsumi Tanaka
IR39	Topic Aspect Analysis for Multi-Document Summarization	Dingding Wang, Shen Chao, Tao Li
IR40	Finding Unusual Review Patterns Using Unexpected Rules	Nitin Jindal, Bing Liu, Ee-Peng Lim
IR41	Visual-Semantic Graphs: Using Queries to Reduce the Semantic Gap in Web Image Retrieval	Barbara Poblete, Benjamin Bustos, Marcelo Mendoza, Juan Barrios
IR42	Novel Local Features with Hybrid Sampling Technique for Image Retrieval	Leszek Kaliciak, Dawei Song, Nirmalie Wiratunga, Jeff Pan
IR43	Expected Browsing Utility for Web Search Evaluation	Emine Yilmaz, Milad Shokouhi, Nick Craswell, Stephen Robertson
IR44	Community-based Topic Modeling for Social Tagging	Daifeng Li, Bing He, ding ying, Jie Tang, Juanzi Li, Zheng Qin, Tianxi Dong, Cassidy Sugimoto
IR45	Discriminative Factored Prior Model for Personalized Content-Based Recommendation	Lanbo Zhang, Yi Zhang
IR46	Fast Query Expansion Using Approximations of Relevance Models	Marc-Allen Cartright, James Allan, Victor Lavrenko, Andrew McGregor
IR47	Mining Rules to Explain Activities in Videos	Florez Omar, Curtis Dyreson
IR48	Online Stratified Sampling: Evaluating Classifiers at Web Scale	Paul Bennett, Vitor Carvalho

IR49	Routing Questions to Appropriate Answerers in Community Question Answering Services	Baichuan Li, Irwin King
IR50	Learning to Rank with Groups	Yuan Lin, Hongfei Lin, Zheng Ye, Song Jin, Xiaoling Sun
IR51	Optimizing Unified Loss for Web Ranking Specialization	Li Fan, Xin Li, Jiang Bian, Zhaohui Zheng
IR52	Hypergraph-based Multilevel Matrix Approximation for Text Information Retrieval	Haw-ren Fang, Yousef Saad
IR53	A Peer-Selection Algorithm for Information Retrieval	Yosi Mass, Yehoshua Sagiv, Michal Shmueli-Scheuer
IR54	Exploring Domain-specific Term Weight in Archived Question Search	Zhao-Yan Ming, Tat-Seng Chua, Gao Cong
IR55	Multi-Information Fusion for Uncertain Semantic Representations of Videos	Bo Lu
IR56	A Topical Link Model for Community Discovery in Textual Interaction Graph	Guoqing Zheng, Jinwen Guo, Lichun Yang, Shengliang Xu, Shenghua Bao, Zhong Su, Dingyi Han, Yong Yu

KM Track

KM1	Taxonomic Clustering of Web Service for Efficient Discovery	Sourish Dasgupta, Satish Bhat, Yugyung Lee
KM2	Active Learning in Parallel Universes	Nicolas Cebron, Michael Berthold
KM3	TAGME: on-the-fly annotation of short text fragments (by Wikipedia entities)	Paolo Ferragina, Ugo Scaiella
KM4	Outlier Ranking in Relevant Subspace Projections	Emmanuel Müller, Matthias Schiffer, Thomas Seidl
KM5	Understanding Retweeting Behaviors in Social Networks	Zi Yang, Jingyi Guo, Jie Tang, keke cai, Juanzi Li
KM6	Mapping Web Pages to Database Records via Link Paths	Weninger Tim, Fabio Fumarola, Jiawei Han, Donato Malerba
KM7	Personalized Recommender System Based on Item Taxonomy and Folksonomy	Huizhi Liang, Yue Xu, Yuefeng Li, Richi Nayak

KM8	Communication Motifs: A Tool to Characterize Social Communications	Qiankun Zhao, Yuan Tian, Qi He, Nuria Oliver, Ruoming Jin, Wang-Chien Lee
KM9	Improving Taxonomies for Large-Scale Hierarchical Classifiers of Web Documents	Kiyoshi Nitta
KM10	PTM: Probabilistic Topic Modeling for Mining Parallel Document Collections	Duo Zhang, Jimeng Sun, ChengXiang Zhai
KM11	Hierarchical Auto-Tagging: Organizing Q&A Knowledge for Everyone	Kyosuke Nishida, Ko Fujimura
KM12	Extracting Structured Information from Wikipedia Articles to Populate Infoboxes	Dustin Lange, Christoph Böhm, Felix Naumann
KM13	Automatic Metadata Extraction from Multilingual Enterprise Content	Melike Sah, Vincent Wade
KM14	Intelligent Sales Forecasting Engine using Genetic Algorithms	Vijayalakshmi Murlidhar, Bernard Menezes, Aniket Divecha, Rohit Menon, Kamal Mehta, Rajesh Ravindran
KM15	Identifying New Categories in Community Question Answering Archives: A Topic Modeling Approach	Yajie Miao
KM16	An Effective Approach for Mining Mobile User Habits	Happia Cao, Tengfei Bao, Qiang Yang, Enhong Chen, Jilei Tian
KM17	Mining Networks with Shared Items	Jun Sese, Mio Seki, Mutsumi Fukuzaki
KM18	Learning Sentiment Classification Model from Labeled Features	Yulan He
KM19	Embedding tolerance relations in formal concept analysis - An application in information fusion	Mehdi Kaytoue, Zainab Assaghir, Sergei O. Kuznetsov, Amedeo Napoli
KM20	Online Learning for Multi-Task Feature Selection	Haiqin Yang, Irwin King, Michael R. Lyu
KM21	Exploiting User Interests for Collaborative Filtering:Interests Expansion via Personalized Ranking	Liu Qi, Enhong Chen, Hui Xiong, Chris Ding
KM22	K-Farthest-Neighbors-Based Concept Boundary Determination for Support Vector Data Description	yanshan xiao, Bo Liu, Longbing Cao

KM23	Relational Feature Engineering of Natural Language Processing	Hamidreza Kobdani, Hinrich Schütze, Andre Burkovski, Stefanie Wiltrud Kessler, Gunther Heidemann
KM24	Transfer Incremental Learning for Pattern Classification	Zhenfeng Zhu, Xingquan Zhu, Yue-Fei Guo, Xiangyang Xue
KM25	Learning Ontology Resolution for Document Representation and its Applications in Text Mining	Lidong Bing, Bai Sun, Shan Jiang, Yan Zhang, Wai Lam
KM26	Supervised Identification and Linking of Concept Mentions to a Domain-Specific Ontology	Gabor Melli, Martin Ester
KM27	Enriching Context in Contextual Advertising	Pavan Kumar GM, Krishna P. Leela, Mehul Parsana, Sachin Garg
KM28	CasJoin: A Cascade Chain for Text Similarity Joins	Xiaoxun Zhang, Zhili Guo, Honglei Guo, Huijia Zhu, Zhong Su
KM29	Learning Naive Bayes Transfer Classifier through Class-wise Test Distribution Estimation	Jeong-Woo Son, Seong-Bae Park and Hyun-Je Song
KM30	TOP-EYE: Top-k Evolving Trajectory Outlier Detection	Yong Ge, Hui Xiong, Zhi-Hua Zhou
KM31	Multi Task Learning on Multiple Related Networks	Prakash Mandayam Comar, Pang-Ning Tan, Anil Kumar Jain
KM32	Building and Evaluating a Semantic Representation for Personal Information	Jinyoung Kim, Anton Bakalov, David Smith, Bruce Croft
KM33	Affinity-Driven Prediction and Ranking of Products in Online Product Review Sites	Hui Li, Sourav S Bhowmick, Aixin Sun
KM34	Topic-Driven Web Search Result Organization by Leveraging Wikipedia Semantic Knowledge	Xianpei Han, Jun Zhao
KM35	Fast Dimension Reduction for Document Classification Based on Imprecise Spectrum Analysis	Guan Hu, Bin Xiao, Jingyu Zhou, Minyi Guo, Tao Yang
KM36	Manifold Ranking with Sink Points for Update Summarization	Pan Du, Jiafeng Guo, Jin Zhang, Xueqi Cheng

KM37	Construction of a Sentimental Word Dictionary	Eduard Dragut, Clement Yu, Prasad Sistla, Weiyi Meng
KM38	Exploiting Novelty, Coverage and Balance for Topic-Focused Multi-Document Summarization	Xuan Li, Yi-dong Shen, Liang Du, Chenyan Xiong
KM39	Context Modeling for Ranking and Tagging Bursty Features in Text Streams	Xin Zhao, Jing Jiang, Jing He, Xiaoming Li, Hongfei Yan, Shan Dongdong
KM40	Argumentation versus Aggregation of Trust Evidence: an Experimental Comparison	Pierpaolo Dondio
KM41	Visualization and Clustering of Crowd Video Content in MPCA Subspace	Haiping Lu, How-Lung Eng, Myo Thida, Konstantinos Plataniotis
KM42	ANITA: A Narrative Interpretation of Taxonomies for their Adaptation to Text Collections	Mario Cataldi, K. Selcuk Candan, Maria Luisa Sapino
KM43	Yes We Can - Simplex Volume Maximization for Descriptive Web-Scale Matrix Factorization	Christian Thurau, Kristian Kersting, Christian Bauckhage
KM44	Incorporating Terminology Evolution for Query Translation in Text Retrieval with Association Rules	Kaluarachchi Amal, Aparna Varde
KM45	Exploiting Co-Occurrence and Information Quality Metrics to Recommend Tags in Web 2.0 Applications	Fabiano Belem, Eder Martins, Jussara Almeida, Marcos Andre Goncalves, Gisele Pappa
KM46	Elusive Vandalism Detection at Wikipedia: A Text Stability-based Approach	Qinyi Wu, Danesh Irani, Albert Larnder, Lakshmish Ramaswamy
KM47	Feature Subspace Transformations for Enhancing K-Means Clustering	Anirban Chatterjee, Sanjukta Bhowmick, Padma Raghavan
KM48	On Bootstrapping Recommender Systems	Yehuda Koren, Nadav Golbandi, Ronny Lempel
KM49	Using Wikipedia Categories for Compact Representations of Chemical Documents	Benjamin Köhncke, Wolf-Tilo Balke
KM50	Efficient Wikipedia-Based Semantic Interpreter by Exploiting Top-k Processing	Jong Wook Kim, Ashwin Kashyap, Dekai Li, Sandilya Bhamidipati
KM51	A study of rumor control strategies on social networks	Rudra Tripathy, Amitabha Bagchi, Sameep Mehta

KM52	Domain-Independent Entity Coreference in RDF Graphs	Dezhao Song, Jeff Heflin
KM53	Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews	Samaneh Moghaddam, Martin Ester
KM54	Combining Link and Content for Collective Active Learning	Lixin Shi, Yuhang Zhao, Jie Tang
KM55	Classifying Sentiment in Microblogs: Is Brevity an Advantage?	Bermingham Adam, Alan Smeaton
KM56	Transient Crowd Discovery on the Real-Time Social Web	Krishna Kamath, James Caverlee
KM57	Discovery of Numerous Specific Topics via Term Co-Occurrence Analysis	Omid Madani, Jiye Yu
KM58	Digging for Knowledge with Information Extraction: A Case Study on Human Gene-Disease Associations	Markus Bundschus, Anna Bauer-Mehren, Volker Tresp, Hans-Peter Kriegel
KM59	Query Log based Personalization using Topic Models	Mark Carman, Morgan Harvey, Mark Baillie, Fabio Crestani
KM60	Choosing Your Own Adventure: Automatic Taxonomy Generation to Permit Many Paths	Xiaoguang Qi, Dawei Yin, Zhenzhen Xue, Brian Davison
KM61	Robust Prediction from Multiple Heterogeneous Data Sources with Partial Information	Mohammad Aziz, Chandan Reddy
KM62	Collaboration Analytics: Mining Work Patterns from Collaboration Activities	Qihua Wang, Hongxia Jin, Yan Liu
KM63	Adapting Cost-sensitive Learning for Reject Option	Jun Du, Eileen Ni, Charles Ling
KM64	SKIF: A Data Imputation Framework for Concept Drifting Data Streams	Zhang Peng, Xingquan Zhu, Jianlong Tan, Li Guo
KM65	Detecting controversial events in Twitter	Ana-Maria Popescu, Pennacchiotti Marco
KM66	Topic Detection and Organization of Mobile Text Messages	Ye Tian, Wendong Wang, Xueli Wang, Jinghai Rao, Canfeng Chen, Jian Ma
KM67	Unsupervised Public Health Event Detection for Epidemic Intelligence	Marco Fisichella, Avaré Stewart , Kerstin Denecke, Wolfgang Nejdl

KM68	Pattern Based Keyword Extraction for Contextual Advertising	Kushal Dave, Dr.Vasudev Varma
KM69	Mixture Model Label Propagation	Mingmin Chi
KM70	Regularization and Feature Selection for Networked Features	Hongliang Fei, Brian Quanz, Luke Huan

Industry Track

IND1	BagBoo: A Scalable Hybrid Model Based on Bagging and Boosting	Alexey Gorodilov, Cliff Brunk, Dmitry Pavlov
IND2	Exploiting Sequential Relationships for Familial Classification	Lee Jensen, James Shanahan
IND3	Massive Structured Data Management Solution	Ullas Nambiar, Himanshu Gupta, rajeev gupta, Mukesh Mohania
IND4	Ranking of Evolving Stories Through Meta-Aggregation	Juozas Gordevicius, Francisco J. Estrada, Hyun Chul Lee, Periklis Andritsos, Johann Gamper
IND5	Injecting Domain Knowledge into a Granular Database Engine: A Position Paper	Dominik Slezak, Graham Toppin
IND6	Experiences with Using SVM-based Learning for Multi-Objective Ranking	Linh Nguyen, Wai Gen Yee, Ophir Frieder
IND7	Learning to Blend Rankings: a monotonic transformation to blend rankings from multiple domains	Zhenzhen Kou, Yi Chang, Zhaohui Zheng, Hongyuan Zha

CIKM'10 Workshops and Tutorials at a Glance

October 26-30, 2010: On-Site Registration *[Second Floor Lobby]*

Tuesday, October 26, 2010

Morning	**AND** *[Room: Confed 5]* Fourth Workshop on Analytics for Noisy Unstructured Text Data	**BooksOnline** *[Room: British Columbia]* Workshop on Research Advances in Large Digital Book Repositories and Complementary Media	**DTMBIO** *[Room: Confed 6]* ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics	**PaIR** *[Room: New Brunswick]* 3rd International Workshop on Patent Information Retrieval	**Tutorial 1** *[Room: Algonquin]* Harvesting Knowledge from Web Data and Text	**Tutorial 2** *[Room: York]* Online Advertising Business Models, Technologies and Issues
Afternoon	**AND** *[Room: Confed 5]* Fourth Workshop on Analytics for Noisy Unstructured Text Data	**BooksOnline** *[Room: British Columbia]* Workshop on Research Advances in Large Digital Book Repositories and Complementary Media	**DTMBIO** *[Room: Confed 6]* ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics	**PaIR** *[Room: New Brunswick]* 3rd International Workshop on Patient Information Retrieval	**Tutorial 3** *[Room: Algonquin]* Real World Text Mining	**Tutorial 4** *[Room: York]* Mobility Data: Modeling, Management, and Understanding

Saturday, October 30, 2010

Morning	**CloudDB** *[Room: Confed 3]* Second International Workshop on Cloud Data Management	**DOLAP** *[Room: Algonquin]* AMC 13th International Workshop on Data Warehousing and OLAP	**ESAIR** *[Room: Confed 5]* Third Workshop on Exploiting Semantic Annotations in Information Retrieval	**PIKM** *[Room: Alberta]* Workshop for Ph.D. Students in Information and Knowledge Management	**SMUC** *[Room: Manitoba]* 2nd International Workshop on Search and Mining User-generated Contents
Afternoon	**CloudDB** *[Room: Confed 3]* Second International Workshop on Cloud Data Management	**DOLAP** *[Room: Algonquin]* AMC 13th International Workshop on Data Warehousing and OLAP	**ESAIR** *[Room: Confed 5]* Third Workshop on Exploiting Semantic Annotations in Information Retrieval	**PIKM** *[Room: Alberta]* Workshop for Ph.D. Students in Information and Knowledge Management	**SMUC** *[Room: Manitoba]* 2nd International Workshop on Search and Mining User-generated Contents

Workshop Programs

Fourth Workshop on Analytics for Noisy Unstructured Text Data (AND 2010)

Tuesday, October 26, 2010 (Room: Confederation 5)

Workshop Organizers:

- Roberto Basili, *University of Rome, Italy*
- Daniel Lopresti, *Lehigh University, United States of America*
- Christoph Ringlstetter, *University of Munich, Germany*
- Shourya Roy, *Xerox India Innovation Hub, India*
- L. Venkata Subramaniam, *IBM Research, India*
- Klaus U. Schulz, *University of Munich, Germany*

Program Schedule:

9:00-9:15 **Welcome**
Daniel Lopresti *(Lehigh University, United States of America)*

9:15-10:00 **Keynote Address**
The Nature of Noise in Linguistic Corpora
Randy Goebel *(University of Alberta, Canada)*

10:00-10:30 **Coffee Break**

10:30-12:30 **Session 1:**
Session Chair: Yosi Mass *(IBM, Haifa, Israel)*

Variant Search and Syntactic Tree Similarity Approach to Retrieve Matching Questions for SMS queries
Akhil Langer, Rohit Banga, Ankush Mittal and Venkata Subramaniam

Clustering Based Approach to Learning Regular Expressions over Large Alphabet for Noisy Unstructured Text
Rohit Babbar and Nidhi Singh

Discovering Users' Topics of Interest on Twitter: A First Look
Matt Michelson and Sofus Macskassy

The effects of learner errors on the development of a collocation detection tool
Yoko Futagi

Tokenizing Micro-Blogging Messages using a Text Classification Approach
Gustavo Laboreiro, Luís Sarmento, Jorge Teixeira and Eugénio Oliveira

Reshaping automatic speech transcripts for robust high-level spoken document analysis
Julien Fayolle, Fabienne Moreau, Christian Raymond and Guillaume Gravier

12:30-14:00 **Lunch**

14:00-15:30 **Session 2:**
Session Chair: Matt Michelson *(University of Southern California, United States of America)*

Statement Map: Reducing Web Information Credibility Noise through Opinion Classification
Koji Murakami, Eric Nichols, Junta Mizuno, Yotaro Watanabe, Shouko Masuda, Hayato Goto, Megumi Ohki, Chitose Sao, Suguru MATSUYOSHI, Kentaro Inui and Yuji Matsumoto

Improving Accuracy of Identifying Clinical Concepts in Noisy un-structured clinical corpus using existing internal redundancy
Pooyan Asgari, Jon Patrick and Negin Motamedi

A Platform for Storing, Visualizing, and Interpreting Collections of Noisy Documents
Bart Lamiroy and Daniel Lopresti

Extracting Person Names from Diverse and Noisy OCR Text
Kevin Seppi and Lee Jensen

15:30-16:00 **Coffee Break**

16:00-17:30 **Session 3:**
Session Chair: Christoph Ringlstetter *(University of Munich, Germany)*

Document: a useful level for facing noisy data
Hervé Déjean and Jean-Luc Meunier

Panel Discussion: Why is it Impossible to handle noisy text with existing techniques: The way forward
Yuji Matsumoto *(Information Science, Nara Institute of Science and Technology, Japan)*
Seamus Ross *(Faculty of Information, University of Toronto, Canada)*
Gareth Jones *(Dublin City University, Ireland)*

17:30-17:45 **Closing**
L. Venkata Subramaniam *(IBM Research, India)*

Third Workshop on Research Advances in Large Digital Book Repositories and Complementary Media (BooksOnline 2010)

Tuesday, October 26, 2010 (Room: British Columbia)

Workshop Organizers:

- Gabriella Kazai *(Microsoft Research, UK)*
- Peter Brisolovsky *(University of Pittsburgh, USA)*

Program Schedule:

9:00-10:00 **Keynote Address**

The Metadata Challenge: Promoting Discovery, Access, and Usability for Online Books
John Mark Ockerbloom *(University of Pennsylvania) – 50 min*

Discussion – *10 min*

10:00-10:30 **Coffee Break**

10:30-11:35 **Session 1: User aspects and design**

The sBook: towards Social and Personalized Learning Experiences
Myriam Ribière, Jérôme Picault and Sylvain Squedin *(Alcatel-Lucent Bell Labs) – 15 min*

Real-Time Document Collaboration Using iPads
Jennifer Pearson *(Swansea University)* and George Buchanan *(City University) – 15 min*

HCI Design Principles for eReaders
Jennifer Pearson *(Swansea University)*, George Buchanan *(City University)* and Harold Thimbleby *(Swansea University) – 25 min*

Discussion – *10 min*

11:35-12:30 **Session 2: Infrastructure aspects**

Ebooks Children Would Want to Read and Engage with
Monica Landoni *(University of Lugano) – 15 min*

Working with First Nations: On-Demand Book Service
Nadia Caidi and Margaret Lam *(University of Toronto) – 15 min*

Implementing New Knowledge Environments: Building Upon Research Foundations to Understand Books and Reading in the Digital Age
Ray Siemens and Julie Meloni *(University of Victoria) – 15 min*

Discussion – *10 min*

14:00-15:00 **Keynote Address**
The Present and Future of Google Books
James Crawford *(Google Books) – 50 min*

Discussion – *10 min*

15:00-15:20 **Poster Session**

Biblioteca de Livros Digitais: The privileged Space of a Transliterate Experience for Children Reading Online
Fernanda Bonacho *(Lisbon Polytechnic Institute)*

Evaluating E-books
Monica Landoni *(University of Lugano)*

A System for the Collaborative Reading of Digital Books with the Partially Sighted
W. Xavier Snelgrove and Ronald M. Baecker *(University of Toronto)*

Enhancing Access To Classic Children's Literature
Claudia Hauff and Dolf Trieschnigg *(University of Twente)*

Can we Trust Majority Rule in Crowdsourced Relevance Evaluation for Digitized Books?
Gabriella Kazai *(Microsoft Research)*

15:20-15:30 **Preparation for break-out sessions**

15:30-16:00 **Coffee Break**

16:00-17:00 **Break-out sessions**
Project proposals for seed funding
Emerging issues and opportunities
Evolving the research agenda

17:00-18:00 **Panel**
Reports from break-out sessions
Feedback from panellists and discussion

Panelists:
James Crawford *(Google Books)*
John Mark Ockerbloom *(University of Pennsylvania)*
Other panelists - TBA

ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics (DTMBIO 2010)

Tuesday, October 26, 2010 (Room: Confederation 6)

Workshop Organizers:

- Doheon Lee *(KAIST, Korea)*
- Hagit Shatkay *(University of Delaware, USA)*
- Min Song *(NJIT, USA)*
- Shamkant Navathe *(Georgia Tech, USA)*

Program Schedule:

9:00-9:15 **Welcome**
Hagit Shatkay *(University of Delaware)*

9:15-10:00 **Keynote Address**
Session Chair: Doheon Lee *(KAIST)*

Information Distance
Ming Li *(University of Waterloo, Canada)*

10:00-10:30 **Coffee Break**

10:30-12:30 **Session 1: Bio-Data/Text Mining I**
Session Chair: Shamkant Navathe *(Georgia Tech)*

Context-specific Gene Regulatory Networks Subdivide Intrinsic Subtypes of Breast Cancer
Sara Nasser, Heather Cunliffe, Michael Black and Seungchan Kim

Automatic classification of sentences for Evidence Based Medicine
Su Nam Kim, David Martinez and Lawrence Cavedon

Processing SPARQL Queries with Regular Expressions in RDF Databases
Jinsoo Lee, Minh-Duc Pham, Jihwan Lee, Wook-Shin Han and Hwanjo Yu

Discovering biological processes and side effects relationship using the process-drug-side effect network
Sejoon Lee, Doheon Lee and Min Song

12:30-14:00 **Lunch**

14:00-15:30 **Session 2: Bio-Data/Text Mining II**
Session Chair: Isabel Segura Bedmar *(Universidad Carlos III de Madrid)*

Word Sense Disambiguation for Event Trigger Word Detection
David Martinez and Timothy Baldwin

Combining syntactic information and domain-specific lexical patterns to extract Drug-Drug Interactions from biomedical texts
Isabel Segura Bedmar, Paloma Martínez Fernández and César de Pablo Sánchez

Deriving a Test Collection for Clinical Information Retrieval from Systematic Reviews
Florian Boudin, Jian-Yun Nie and Martin Dawes

Unsupervised Word Sense Disambiguation in Biomedical Texts with Co-occurrence Network and Graph Kernel
Tae Gil Noh and Seong-Bae Park

15:30-16:00 **Coffee Break**

16:00-16:30 **Session 3: Short Presentation**
Session Chair: Wook-Shin Han *(Kyungpook National University)*

Dynamic Concept Ontology Construction for PubMed Queries
Jinoh Oh, Taehoon Kim, Sun Park, Hwanjo Yu and Wook-Shin Han

Effect of Classifiers in Consensus Feature Ranking for Biomedical Datasets
Shobeir Fakhraei, Hamid Soltanian-Zadeh, Farshad Fotouhi and Kost Elisevich

Recent research for MEDLINE/PubMed: Short Review
Illhoi Yoo and Miroslav Marinov

16:30-17:15 **Session 4: Joint Discussion/Wrap-Up**
Joint Discussion of all Papers and DTMBIO Workshop

Third International Workshop on Patent Information Retrieval (PaIR 2010)

Tuesday, October 26, 2010 (Room: New Brunswick)

Workshop Organizers:

- John Tait *(IRF, Austria)*
- Christopher Harris *(University of Iowa, USA)*
- Mihai Lupu *(IRF, Austria)*
- Katja Mayer *(IRF, Austria)*

Program Schedule:

9:00-9:10 **Welcome**
Mihai Lupu *(IRF)*

9:10-10:10 **Keynote Address**
Session Chair: Mihai Lupu *(IRF)*

Invited Talk by Mary Carman, *CEO, Canada Intellectual Property Office*

10:00-10:30 **Coffee Break**

10:30-12:30 **Session 1**
Session Chair: Mihai Lupu *(IRF)*

Search for Patents Using Treatment and Causal Relationships
Ashwathi Krishnan, Alfonso Cardenas and Derek Springer

Automatic Creation of a Technical Trend Map from Research Papers and Patents
Hidetsugu Nanba, Tomoki Kondo and Toshiyuki Takezawa *(short presentation)*

Aspects and Analysis of Patent Test Collections
Mihai Lupu, Florina Piroi and Allan Hanbury. *(short presentation)*

Unsupervised Learning for Reranking-based Patent Retrieval
Wenhui Liao and Sriharsha Veeramachaneni. *(short presentation)*

Comparison of IPC and USPC Classification Systems in Patent Prior Art Searches
Christopher Harris, Robert Arens and Padmini Srinivasan. *(short presentation)*

PatentsSearcher: A Novel Portal to Search and Explore Patents
Vagelis Hristidis, Eduardo Ruiz, Alejandro Hernández, Fernando Farfán and Ramakrishna Varadarajan. *(short presentation)*

Genre and Domain in Patent Texts
Nelleke Oostdijk, Hans van Halteren, Eva D'hondt and Suzan Verberne.

12:05-12:30 **Poster Session (Room: Nova Scotia)**

12:30-14:00 **Lunch**

14:00-15:00 **Keynote Address**
Session Chair: Mihai Lupu *(IRF)*

Results of User Research for Improving Patent Examination Tools
Marti Hearst, *US Patent and Trademark Office*

15:00-16:00 **Session 2**
Session Chair: Christopher Harris *(Univeristy of Iowa)*

Vector Space Analysis of Swedish Patent Claims with Different Linguistic Indices
Linda Andersson

Preliminary Study into Query Translation for Patent Retrieval
Charles Jochim, Christina Lioma, Hinrich Schütze, Steffen Koch and Thomas Ertl

16:00-16:30 **Coffee Break**

16:30-17:20 **Session 3: Panel Discussion**
Panel Moderator: Christopher Harris *(University of Iowa)*

17:20-17:30 **Conclusion**

Second International Workshop on Cloud Data Management (CloudDB 2010)

Saturday, October 30. 2010 (Room: Confederation 3)

Workshop Organizers:

- General Co-Chairs:
  - Prof. Xiaofeng Meng, *Renmin University of China, China*
  - Dr. Ying Chen, *IBM China Research Lab, China*
- PC Co-Chiars:
  - Dr. Jianliang Xu, *Hong Kong Baptist University, Hong Kong*
  - Dr. Jiaheng Lu, *Renmin University of China, China*
  - Dr. Qiu Jie, *IBM Research Lab, China*

Program Schedule:

9:00-9:15 **Welcome**
Xiaofeng Meng *(Renmin University of China)*

9:15-10:15 **Session 1: Query Processing in Cloud DBs**
Session Chair: Xiaofeng Meng

Adaptive Query Execution for Data Management in the Cloud
Adrian Daniel Popescu, Debabrata Dash, Verena Kantere

ESQP: An Efficient SQL Query Processing for Cloud Data Management
Jing Zhao, Xiangmei Hu, Xiaofeng Meng

10:15-10:30 **Coffee Break**

10:30-12:00 **Session 2: Cloud-based Architecture, Benchmarking and Security**
Session Chair: Jie Qiu

Contract-based Cloud Architecture
Maxim Schnjakin, Rehab Alnemr, Christoph Meinel

Benchmarking Cloud-based Data Management Systems
Yingjie Shi, Xiaofeng Meng, Jing Zhao, Xiangmei Hu, Bingbing Liu, Haiping Wang

Towards a Data-centric View of Cloud Security
Wenchao Zhou, Micah Sherr, William R. Marczak, Zhuoyao Zhang, Tao Tao, Boon Thau Loo, Insup Lee

12:00-14:00 **Lunch**

14:00-15:30 **Session 3: Cloud-based Virtualization and Application**
Session Chair: Jie Qiu

Comparing SQL and MapReduce to compute Naive Bayes in a Single Table Scan

Ordonez, Carlos

Dynamic Data Replication through Virtualization
Sergey Savinov, Khuzaima Daudjee

Towards Bipartile Graph Data Management
Bin Zhao, Weining Qian, Aoying Zhou

15:30-16:00 **Coffee Break**

16:00-17:00 **Joint Discussion/Wrap-up**
Joint Discussion of All Papers and CloudDB Workshop

ACM Thirteenth International Workshop on Data Warehousing and OLAP (DOLAP 2010)

Saturday, October 30, 2010 (Room: Algonquin)

Workshop Organizers:

- Il-Yeol Song *(Drexel University, USA)*
- Carlos Ordonze *(University of Houston, USA)*

Program Schedule:

8:00-8:15 **Workshop Welcome and Introduction**

8:15-9:00 **Keynote Address**
Session Chair: C. Ordonez *(University of Houston)*

Query Processing in Oracle DBMS
Rafi Ahmed *(Oracle Corporation)*

9:00-10:30 **Session 1: Conceptual Modeling and Data Warehouse Design**
Session Chair: A. Cuzzocrea *(University of Calabria)*

Using Ontologies to Discover Fact IDs
Alberto Abello, Oscar Romero - *23 min*

Multiversion Data Warehouse Constraints
Ines Zouari, Faiza Ghozzi, Rafik Bouaziz – *23 min*

A Methodology and Tool for Conceptual Designing a Data Warehouse from Ontology-based Sources
Ladjel Bellatreche, Selma Khouri – *16 min*

Graph Layout Readability in OLAP Diagrams for Conceptual Modeling of Data Warehouses
Jesus Pardillo, Florian Mansmann – *16 min*

Towards the Definition of Spatial Data Warehouses Integrity Constraints with OCL
Kamal Bouil, Sandro Bimonte, Hadj Mahboubi, Francois Pinet – *16 min*

10:30-11:00 **Coffee Break**

11:00-12:30 **Session 2: OLAP Query Processing**
Session Chair: A. Abello *(University Politecnica de Catalynya)*

Towards OLAP Query Reformulation in Peer-to-Peer Data Warehousing
Matteo Golfarelli, Federica Mandreoli, Wilma Penzo, Stefano Rizzi, Elisa Turricchia – *23 min*

Immediate Materialized Views with Outerjoins
Anisoara Nica – *23 min*

R-MESHJOIN for Near-Real-Time Data Warehousing
Muhammad Asif Naeem, Gillian Dobbie, Gerald Weber, Shafiq Alam – *23 min*

Repairing OLAP Queries in Databases with Referential Integrity Errors
Javier Garcia-Garcia, Carlos Ordonez – 16 min

12:30-14:00 **Lunch**

14:00-15:00 **Panel: Relational versus Non-Relational Database Systems for Data Warehousing**
Moderator: Carlos Ordonze *(University of Houston)*

Panel Members:
Yu Xu *(Teradata)*
Florian Waas *(Greenplum)*
Ingmar Weber *(Yahoo! Research)*
Yannis Sismanis *(IBM)*

15:00-15:30 **Coffee Break**

15:30-17:30 **Session 3: New Trends**
Session Chair: E. Zimanyi *(University Libre de Bruxelles)*

Time-HOBI: Indexing Dimension Hierarchies by Means of Hierarchically Organized Bitmaps
Jan Chmiel, Tadeusz Morzy, Robert Wrembel – *23 min*

Exploring Graphics Processing Units as Parallel Coprocessors for Online Aggregation
Tobias Lauer, Amitava Datta, Zurab Khadikov, Christoffer Anselm – *23 min*

Integrating OLAP and Recommender Systems: An Evaluation Perspective
Artus Krohn-Grimberghe, Alexandros Nanopoulos, Lars Schmidt-Thieme – *23 min*

Balancing Accuracy and Privacy of OLAP Aggregations on Data Cubes
Alfredo Cuzzocrea, Domenico Sacca – *16 min*

Context-Aware Generalization for Cube Measures
Yoann Pitarch, Cecile Favre, Anne Laurent, Pascal Poncelet – *16 min*

17:30-18:00 **Open-floor discussion and workshop summary**

Third Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR 2010)

Saturday, October 30, 2010 (Room: Confederation 5)

Workshop Organizers:

- Jaap Kamps *(University of Amsterdam)*
- Jussi Karlgren *(SICS, Stockholm)*
- Ralf Schenkel *(MPI/Saarland University)*

Program Schedule:

9:00-9:30 **Welcome and Feature Rally**

9:30-10:00 **Keynote Address**
Session Chair: Jussi Karlgren

Questions to be Asked & Answered as to NLP's Role in Improving Semantic Annotation
Elizabeth D. Liddy *(Syracuse University)*

10:00-10:30 **Coffee Break**

10:30-11:00 **Keynote Address**
Session Chair: Jaap Kamps

The Surplus Value of Semantic Annotations
Maarten Marx *(University of Amsterdam)*

11:00-12:30 **Boasters and Posters Session *(Room: Confederation 6)***

Applications ***The Crucial Role of Semantic Discovery and Markup in Geo-temporal Search***
Fredric Gey, Noriko Kando and Ray Larson

Semantic Annotations for Digital Investigations
Nikolaos Lagos, Stefania Castellani and Aaron Kaplan

Semantic Annotations in Clinical Documentation
Sumithra Velupillai

Annotation ***Cross language information retrieval based on Concept base and Language Grid***
Pham Huy Anh and Takashi Yukawa

Is Formalizing Events Necessary for Full Exploitation
Antonio Badia

TAGME: On-the-fly Annotation of Short Test Fragments
Paolo Ferragina and Ugo Scaiella, *CIKM poster re-presented at ESAIR*

On the Definition of Patterns for Semantic Annotation
Monica Marrero, Julián Urbano, Jorge Morato and Sonia Sánchez-Cuadrado

Creating Software Models with Semantic Annotation
Walter Tichy, Sven Koerner and Mathias Landhäußer

Tagging for Improved Semantic Interpretation of XML Documents
Vicente Palacios, Juan Lloréns, Sonia Sánchez-Cuadrado and Monica Marrero

Aggregation

SQR: A Semantic Query Rating Scheme
Hany Azzam and Thomas Roelleke

Application of Semantic Annotations to Predicting Users' Demographics
Blaz Fortuna, Dunja Mladenić and Marko Grobelnik

Generating Document Summaries from User Annotations
Karen Shiells, Omar Alonso and Ho John Lee

Search by Strategy
Arjen de Vries, Wouter Alink and Roberto Cornacchia

Searchers & Queries

A Tool for Ontology-Editing and Ontology-Based Information Exploration
Feza Baskaya, Jaana Kekäläinen and Kalervo Järvelin

A Semantic Annotation Framework for Retrieving and Analyzing Observational Datasets
Shawn Bowers, Huiping Cao, Mark Schildhauer, Matt Jones, Ben Leinfelder and Margaret O'Brien

Modeling Betweenness for Question Answering
Brandeis Marshall

Exploiting Hierarchical Tags for Context-awareness
Alan Said and Ernesto W. De Luca

12:30-14:00 **Lunch**

14:00-15:30 **Break-out Sessions *(Rooms: Confederation 5 and Confederation 6)***
Focus groups discussing particular aspects in parallel

15:30-16:00 **Coffee Break**

16:00-17:00 **Reports from the Break-Out Groups + Workshop Wrap-up**

ACM Workshop for Ph.D. Students in Information and Knowledge Management (PIKM 2010)

Saturday, October 30, 2010 (Room: Alberta)

Working Co-Chairs:

- Anisoara Nica, *Sybase iAnywhere, Waterloo, Canada*
- Aparna S. Varde, *Montclair State University, USA*

Program Schedule:

9:00-9:15 **Welcome**
Aparna S. Varde, *Montclair State University, USA*

9:15-10:00 **Keynote Address**
Session Chair: Anisoara Nica, *Sybase iAnywhere, Waterloo, Canada*
Schema Discovery: Prof. Renee Miller, *University of Toronto, Canada*

10:00-10:30 **Coffee Break and Poster Session**

10:30-12:30 **Session 1: Database Track (DB)**
Session Chair: Aparna S. Varde, *Montclair State University, USA*

Annotations: Dynamic Semantics in Stream Processing
Juan Amiguet, Andreas Wombacher, Tim E. Klifman

Adaptive Query Processing in Data Stream Management Systems Under Limited Memory Resources
Fatima Farag, Moustafa Hammad, Reda Alhaji

Efficient Algorithms Based on Relational Queries to Mine Frequent Graphs
Walter Garcia, Carlos Ordonez, Kai Zhao, Ping Chen

Quality Factory and Quality Notification Service in Data Warehouse
Yan Li, Kweku-Muata Osei-Bryson

12:30-14:00 **Lunch**

14:00-15:30 **Session 2: Knowledge Management Track (KM)**
Session Chair: Fabian M. Suchanek, *INRIA, France*

Entity Classification by Bag of Wikipedia Articles
Tomas Kliegr

A Decision Support System for Green Data Centers
Michael J. Pawlish, Aparna S. Varde

Leveraging Natural Language Processing of Clinical Narratives for Phenotype Modeling
Preethi Raghavan, Albert M. Lai

15:30-16:00 **Coffee Break and Poster Session**

16:00-17:00 **Session 3: Information Retrieval Track (IR)**
Session Chair: Anisoara Nica, *Sysbase iAnywhere, Waterloo, Canada*

Research Proposal for Distributed Deep Web Search
Kien-Tsoi T.E. Tjin-Kam-Jet

From Exploratory Search to Web Search and Back
Roberto Mirizzi, Tommaso Di Noia

17:00-18:00 **Session 4: Poster Session**

Identifying the Challenges for Optimizing the Process to Achieving Reproducibility of Results
Mohammad Rezwanul Huq, Andreas Wombacher, Peter M.G. Apers

Probabilistic Ranking for Relational Databases based on Correlations
Jaehui Park, Sang-Goo Lee

Toward the Design of a Methodology to Predict Relevance Through Multiple Sources of Evidence
Emanuele Di Buccio, Massimo Melucci

A Secure Pervasive Health Care System Using Location Dependent Unicast Key Generation Scheme
Debargh Acharya, Vijay Kumar

A Multi-Functional Architecture Addressing Workflow and Service Challenges Using Provenance Data
Mahsa Naseri, Simone A. Ludwig

Modeling a Trust Cloud Context
Sean Thorpe

Second International Workshop on Search and Mining User-Generated Contents (SMUC 2010)

Saturday, October 30, 2010 (Room: Manitoba)

Workshop Organizers:

- Jose Carlos Cortizo *(BrainSins, Spain)*
- Francisco Manuel Carrero *(BrainSins, Spain)*
- Ivan Cantador *(Autonomous University of Madrid, Spain)*
- Jose Troyano *(University of Seville, Spain)*
- Paolo Rosso *(Technical University of Valencia, Spain)*

Program Schedule:

8:00-8:15 **Welcome**
Jose Carlos *(BrainSins)*

8:15-9:15 **Keynote Address**
Session Chair: Paolo Rosso *(Technical University of Valencia)*

Beyond Sentiments – Opinion Mining in the Real World
Bing Liu *(University of Illinois at Chicago)*

9:15-10:00 **Poster Session**
Session Chair: Ivan Cantador *(Autonomous University of Madrid)*

Mining Social Tags to Predict Mashup Patterns
Khaled Goarany, Gregory Kulczycki and M.Brian Blake

A Weighted Tag Similarity Measure Based on a Collaborative Weight Model
Srinivas Gokavarapu, Niket Tandon and Vasudeva Varma

How to interpret the helpfulness of online product reviews: bridging the needs between customers and designers
Jian Jin, Ying Liu and Ee-Peng Lim

Web-based Statistical Fact Checking of Textual Documents
Amr Magdy and Nayer Wanas

On the Difficulty of Clustering Company Tweets
Fernando Perez, David Pinto, John Cardiff and Paolo Rosso

Cross-media Impact on Twitter in Japan
Sayaka Akioka, Norikazu Kato, Yoichi Muraoka and Hayato Yamana

Extracting Emotion Topics from Blog Sentences – Use of Voting from Multi-Engine Supervised Classifiers
Dipankar Das and Sivaji Bandyopadhyay

10:00-10:30 **Coffee Break**

10:30-12:30 **Session 1: Webpages and Product Reviews**
Session Chair: Jose A. Troyano *(University of Seville)*

Exploiting Tag and Word Correlations for Improved Webpage Clustering
Anusua Trivedi, Piyush Rai and Scott DuVall

A Knowledge-Rich Approach to Feature-Based Opinion Extraction from Product Reviews
Fermín L. Cruz Mata, José A. Troyano Jiménez, Fernando Enríquez de Salamanca Ros, F. Javier Ortega Rodríguez and Carlos García Vallejo

A Formal Study of classification Techniques on Entity Discovery and their Application to Opinion Mining
Shadi Banitaan, Saeed Salem, Wei Jin and Ibrahim Aljarah

Spam Detection with a Content-based Random-walk Algorithm
F. Javier Ortega, Craig Macdonald, José A. Troyano and Fermín Cruz

Exploiting Web Reviews for Generating Customer Service Surveys
Suke Li and Zhong Chen

12:30-14:00 **Lunch**

14:00-15:30 **Session 2: Twitter and Wikipedia**
Session Chair: Francisco M. Carrero *(BrainSins)*

Entity-Relationship Queries over Wikipedia
Xiaonan Li, Chengkai Li and Cong Yu

Classifying Latent User Attributes in Twitter
Delip Rao, David Yarowsky, Abhishek Srivats and Manaswi Gupta

Characterization of the Twitter @replies Network: are user ties social or topical?
Daniel Sousa, Luís Sarmento and Eduarda Mendes Rodrigues

15:30-16:00 **Coffee Break**

16:00-17:30 **Industry Panel**
Session Chair: Jose Carlos Cortizo *(BrainSins)*

Panelists:
Nayer Wanas *(Nile Innovations, Egypt)*
Masoud Makrehchi *(Thomson Reuters, USA)*
Jérôme Picault *(Alcatel-Lucent Bell Labs)*
Haggai Rotman *(IBM)*

CIKM'10 Tutorials

Tutorial 1: Harvesting Knowledge from Web Data and Text

Tuesday, October 26, 2010 ***(Room: Algonquin)***

Tutorial Organizers:
Hady W. Lauw, Ralf Schenkel, Fabian Suchanek, Martin Theobald, and Gerhard Weikum

Time: ***9:00 – 12:30***

Abstract:

The Web bears the potential of being the world's greatest encyclopedic source, but we are far from fully exploiting this potential. Valuable scientific and cultural content is interspersed with a huge amount of noisy, lowquality, unstructured text and media. The proliferation of knowledge-sharing communities like Wikipedia and the advances in automated information extraction from Web pages give rise to an unprecedented opportunity: Can we systematically harvest facts from the Web and compile them into a comprehensive machine-readable knowledge base? Such a knowledge base would contain not only the world's entities, but also their semantic properties, and their relationships with each other. Imagine a "Structured Wikipedia" that has the same scale and richness as Wikipedia itself, but offers a precise and concise representation of knowledge, e.g., in the RDF format. This would enable expressive and highly precise querying, e.g., in the SPARQL language (or appropriate extensions), with additional capabilities for informative ranking of query results. The benefits from solving the above challenge would be enormous. Potential applications include 1) a formalizedmachine-readable encyclopedia that can be queriedwith high precision like a semantic database; 2) a key asset for disambiguating entities by supporting fast and accurate mappings of textual phrases onto named entities in the knowledge base; 3) an enabler for entity-relationship-oriented semantic search on the Web, for detecting entities and relations in Web pages and reasoning about them in expressive (probabilistic) logics; 4) a backbone for natural-language question answering that would aid in dealing with entities and their relationships in answering who/where/when/ etc. questions; 5) a key asset for machine translation (e.g., English to German) and interpretation of spoken dialogs, where world knowledge provides essential context for disambiguation; 6) a catalyst for acquisition of further knowledge and largely automatedmaintenance and growth of the knowledge base. While these application areas cover a broad, partly AI-flavored ground, the most notable one from a database perspective is semantic search: finally bringing DB methodology to Web search! For example, users (or tools on behalf of users) would be able to formulate queries about succulents that grow both in Africa and America, politicians who are also scientists or are married to singers, or flu medication that can be taken by people with high blood pressure. The search engine would return precise and concise answers: lists of entities or entity pairs (depending on the question structure), for example, Angela Merkel, Benjamin Franklin, etc., or Nicolas Sarkozy for the questions about scientists. This would be a quantum leap over today's search where answers are embedded if not buried in lots of result pages, and the human users would have to read them to extract entities and connect them to other entities. In this sense, the envisioned large-scale knowledge harvesting fromWeb sources may also be viewed as machine reading.

Tutorial 2: Online Advertising Business Models, Technologies and Issues

Tuesday, October 26, 2010 ***(Room: York)***

Tutorial Organizers:
Deepak K Agarwal, James G. Shanahan

Time: ***9:00 – 12:30***

Abstract:

Over the past 15 years online advertising, a $65 billion industry worldwide in 2008, has been pivotal to the success of the world wide web. This success has arisen largely from the transformation of the advertising industry from a low-tech, human intensive, "Mad Men" way of doing work (that were common place for much of the 20th century and the early days of online advertising) to highly optimized, mathematical, computer-centric processes (some of which have been adapted from Wall Street) that form the backbone of many current online advertising systems. This half-day tutorial (3-4 hours) focuses on helping researchers and developers attending CIKM acquire new skills as well providing an account of the latest developments in this diverse and relatively new scientific discipline of online advertising. In doing so it will provide a clear and detailed overview of the technologies and business models that are transforming the field of (online) advertising along the following themes: business models; market size and scope; forward and spot markets; machine learning and statistical technologies; economic models; new directions such as social advertising and behavioral targeting; challenges and open issues that face online advertising. Participants will learn about: • the primary business models that make online advertising • the online advertising markets, their revenues and sizes • forward and spot markets • linear programming, quadratic programming, Markowitz model • auction models, game theory and how it can be used to analyze online auctions • metrics and evaluation practices (online evaluation through designed experiments like AB tests, Factorial designs; offline evaluation through logged data) • Modern collaborative filtering methods • Bayesian models, online learning and multi-armed bandit schemes. • learning to rank • predict ad quality/click-thru-rates • automatically targeting ads both in sponsored search and in contextual advertising • understand the specific issues that face online advertising such as privacy, deception and fraud (in particular, click fraud, the spam of online advertising) and current approaches • how advertising is being leveraged in Web 2.0 applications such as social networks, and video/photo-sharing • new directions: behavioral targeting, social advertising, data exchanges • the open research issues that face the field of online advertising.

Tutorial 3: Real World Text Mining

Tuesday, October 26, 2010 ***(Room: Algonquin)***

Tutorial Organizers:
Ronen Feldman and Lyle Ungar

Time: ***14:00 – 17:30***

Abstract:

The proliferation of documents available on the Web and on corporate intranets is driving a new wave of text mining research and application. Earlier research addressed extraction of information from relatively small collections of well-structured documents such as newswire or scientific publications. Text mining from the other corpora such as the web requires new techniques drawn from data mining, machine learning, NLP and IR. Text mining requires preprocessing document collections (text categorization, information extraction, term extraction), storage of the intermediate representations, analysis of these intermediate representations (distribution analysis, clustering, trend analysis, association rules, etc.), and visualization of the results. In this tutorial we will present the algorithms and methods used to build text mining systems. The tutorial will cover the state of the art in this rapidly growing area of research, including recent advances in unsupervised methods for extracting facts from text and methods used for web-scale mining. We will also present several real world applications of text mining. Special emphasis will be given to lessons learned from years of experience in developing real world text mining systems, including how to handle informal texts such as blogs and user reviews, infer sentiment towards various objects and how to build scalable systems.

Tutorial 4: Mobility Data: Modeling, Management, and Understanding

Tuesday, October 26, 2010 ***(Room: York)***

Tutorial Organizers:
Stefano Spaccapietra, Esteban Zimányi, Chiara Renso

Time: ***14:00 – 17:30***

Abstract:

Applications using mobility data are blooming in every economic area. Many research efforts have been devoted to developing concepts, models, theories and tools to properly apprehend mobility data and make it manageable for the benefit of these applications. This proposal is authored by experts that have significantly contributed to the field in complementary ways. The tutorial aims at discussing different facets of mobility data, from spatio-temporal data modeling to data aggregation and warehousing and to data analysis, and from basic definitions to state-of-the-art concepts and techniques. One peculiarity of the tutorial is to use a consistent framework that facilitates understanding of all different facets. The learning objective is to enable attendants to understand the many issues at hand and identify the research perspectives they may be willing to investigate to contribute to the domain.

Restaurant Information

Richtree Market
181 Bay Street
Toronto, ON
416.863.0108
http://www.richtree.ca/

European and Canadian
Under $10 per entree

Duke of Devon
66 Wellington Street West
Toronto, ON
416.642.3853
http://devon.thedukepubs.ca/

Pub Food
$15-25 per entree

Canoe Restaurant & Bar
66 Wellington Street West
Toronto, ON
416.364.0054
http://oliverbonacini.com/

Continental and Canadian
$25 on up per entree

Bymark
66 Wellington Street West
Toronto, ON
416.777.1144
http://www.bymarkdowntown.com/

International and Continental
$25 on up per entree

Ki Resturant and Lounge
181 Bay Street
Toronto, ON
416.308.5888
http://www.kijapanese.com

Japanese and Fusion
$15-25 per entree

Loose Moose Tap & Grill
146 Front Street West
Toronto, ON
416.977.8840
http://www.theloosemoose.ca/

American and Barbecue
$10-15 per entree

Bardi's Steak House
56 York Street
Toronto, ON
416.366.9211
http://www.bardis.com/

Steakhouse and Seafood
$25 on up per entree

Jimmy the Greek
100 Wellington Street West
Toronto, ON
416.861.9028
http://www.jimmythegreek.com/

Greek
Under $10 per entree

Tamarind the Indian Kitchen
161 Bay Street
Toronto, ON
416.850.2929
http://www.tamarindkitchen.com/

Indian
Under $10 per entree

Mr. Sub
100 Wellington Street West
Toronto, ON
416.955.8899
http://www.mrsub.ca/

Sandwiches/Subs
$15-25 per entree

Harvey's
65 Front Street West
Toronto, ON
416.364.6185
http://www.harveys.ca/

Fast Food
Under $10 per entree

Tim Hortons
55 York Street
Toronto, ON
416.363.1055
http://www.timhortons.com/

Coffee/Tea and Donuts
Under $10 per entree

Opa! Souvlaki of Greece
200 Bay Street
Toronto, ON
416.368.9600
http://www.opasouvlaki.ca

Greek
Under $10 per entree

Quiznos
100 Wellington Street West
Toronto, ON
416.304.1055
http://www.quiznos.ca/

Sandwiches/Subs
Under $10 per entree

The Bagel Stop
161 Bay Street
Toronto, ON
416.955.9095
http://www.thebagelstop.com/

Bagels and Sandwiches/Subs
Under $10 per entree

Union Station also has a wide variaty dining out selections starting from under $10 per entree

Liberty Grand Banquet Dinner Information

Standing majestically at Toronto's downtown lakefront, the Liberty Grand is the crown jewel of Exhibition Place. Originally constructed in 1926, this building has an impressive heritage. Not only does it recapture the original features but they restored the 1920's splendor and status as one of the city's finest.

Gracefully blending the exuberant Beaux Arts monumentality of the exterior with the classical interior finishes, you will enjoy the epitome of luxury and elegance. Dedicated to the highest standard of hospitality, the Liberty Grand offers the ultimate experience in customer service. This magnificent complex exudes an ambiance of style and refined beauty unparalleled in Toronto. You will be surrounded by an atmosphere of grandeur and uncompromising quality, as well as a most spectacular lakefront view.

25 British Columbia Road, Exhibition Place
Toronto, ON
Tel: 416.542.3789
Web: http://www.libertygroup.com

Transportation will be provided to conference participants. On October 28th, conference participants are asked to be at the Royal York Hotel in the *Front Street enterance at 5:45pm*. The banquet *will conclude at 9:30pm*, at which time participants are asked to be at the *Drive Entrance*. We ask everyone to be timely as delays will result in the conference organizers enduring additional fees.

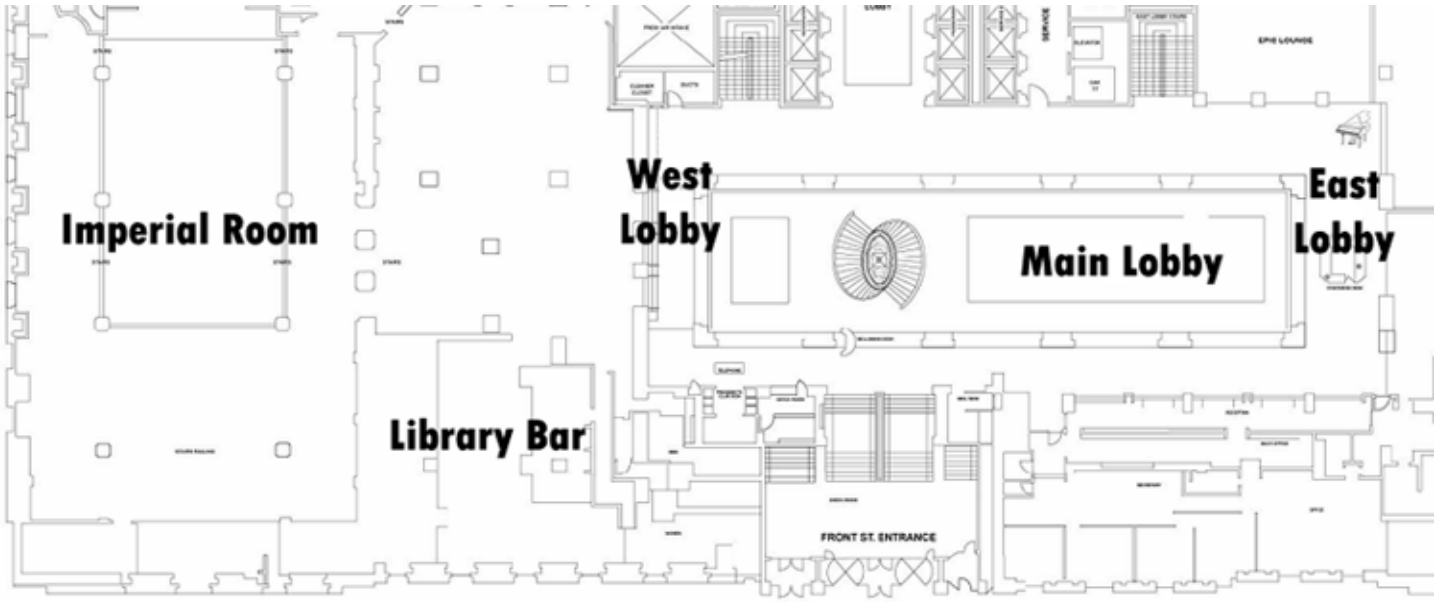
If you wish to take the TTC instead, below are directions:

Directions from Royal York Hotel on October 28th, 2010

1. Leave at 5:47pm from King Street West at York Street
2. Using these routes
    a. Take bus 508 Lake Shore towards Long Branch
    b. Take bus 29D Dufferin towards Exhibition (Princess' Gates)
3. Arriving at 6:11pm Saskatchewan Road at Dufferin Street

For other TTC schedules, please visit *http://www3.ttc.ca/Trip_planner/index.jsp* to use their trip planner.

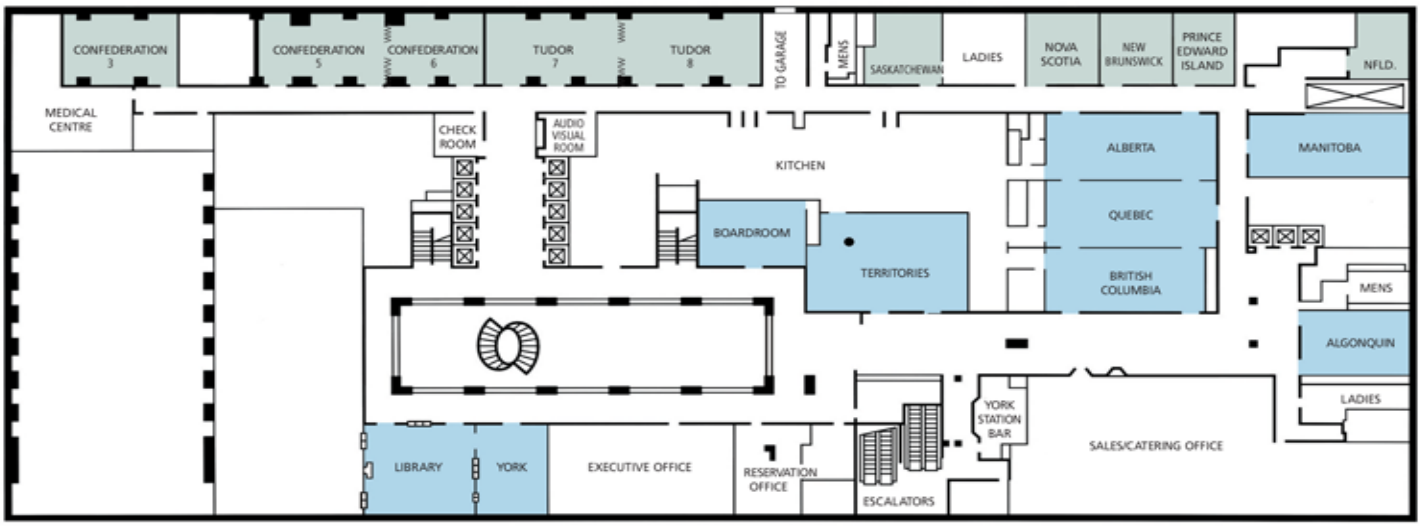
Royal York Hotel Maps



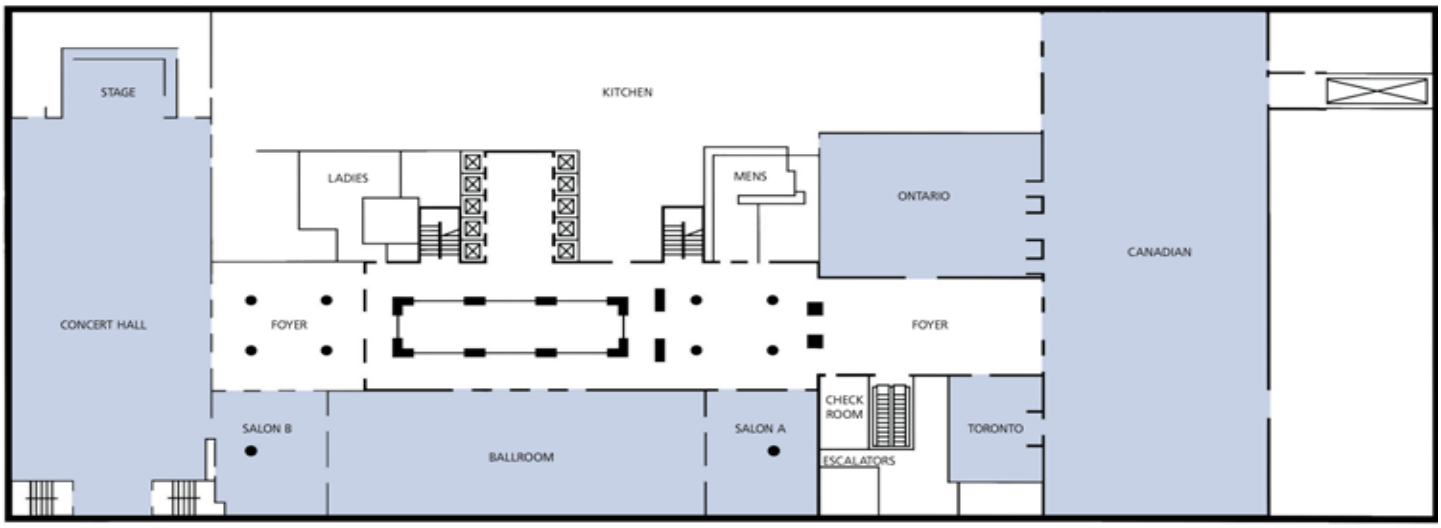
THE Fairmont ROYAL YORK

RYH
LOBBY LEVEL
EFFECTIVE JULY 2002

MAIN MEZZANINE FLOOR



CONVENTION FLOOR



Special Thanks to Our Sponsors

Association for Computing Machinery

KDD.org

SIGIR Special Interest Group on Information Retrieval

sigweb

UMBC AN HONORS UNIVERSITY IN MARYLAND

Яндекс Yandex

DossierView Discover. Distill. Discuss.

Google

YAHOO! LABS

IBM Research

ebay

INFORMATION RETRIEVAL FACILITY IRF QUAERO QUOD REPERIO

MITACS

INFOBRIGHT

NOKIA Connecting People

SOSO 搜搜

Microsoft Research

GenieKnows.com

CAMBRIDGE UNIVERSITY PRESS

YORK UNIVERSITÉ UNIVERSITY

redefine THE POSSIBLE.