

# CIKM'10 Program at a Glance

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

# CIKM'10 Main Conference Program

Wednesday, October 27, 2010

## Conference Opening (8:10-8:30)

**Room:** Imperial

**Welcome:** TBA

**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 Szumlanski 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 Content, Xinying Song, Jing Liu, Yunbo Cao and Chin-Yew Lin
- ◆ An Efficient Algorithm for Mining Time Interval-based Patterns in Large Database, Yi-cheng Chen, 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 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 Kensch 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: Dominik Slezak**

**Room: Tudor 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: TBA**

**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 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 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)**

**Session – 5A (October 28, 13:30 - 15:30)**

**KM Track: Temporal, Spatial and Stream Data Mining**

**Chair: TBA**

**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 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 Faruque, 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(1) (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 Nørkvåg
- ◆ 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)**

**Chair: 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, Ryan 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 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, 3<sup>rd</sup> 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, Andrius 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, Independent 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 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 Bhattacharjee and Vikas Garg

## CIKM'10 Workshops and Tutorials at a Glance

On-site Registration October 26-30, 8:00-18:00 [Second Floor Lobby]						
Tuesday, October 26, 2010						
Morning	<b>AND</b> [Room: Confed 5] Fourth Workshop on Analytics for Noisy Unstructured Text Data	<b>BooksOnline</b> [Room: British Columbia] Workshop on Research Advances in Large Digital Book Repositories and Complementary Media	<b>DTMBIO</b> [Room: Confed 6] ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics	<b>PaIR</b> [Room: New Brunswick] 3rd International Workshop on Patent Information Retrieval	<b>Tutorial 1</b> [Room: Algonquin] Harvesting Knowledge from Web Data and Text	<b>Tutorial 2</b> [Room: York] Online Advertising Business Models, Technologies and Issues
Afternoon	<b>AND</b> [Room: Confed 5] Fourth Workshop on Analytics for Noisy Unstructured Text Data	<b>BooksOnline</b> [Room: British Columbia] Workshop on Research Advances in Large Digital Book Repositories and Complementary Media	<b>DTMBIO</b> [Room: Confed 6] ACM Fourth International Workshop on Data and Text Mining in Biomedical Informatics	<b>PaIR</b> [Room: New Brunswick] 3rd International Workshop on Patent Information Retrieval	<b>Tutorial 3</b> [Room: Algonquin] Real World Text Mining	<b>Tutorial 4</b> [Room: York] Mobility Data: Modeling, Management, and Understanding
Saturday, October 30, 2010						
Morning	<b>CloudDB</b> [Room: Confed 3] Second International Workshop on Cloud Data Management	<b>DOLAP</b> [Room: Algonquin] ACM 13th International Workshop on Data Warehousing and OLAP	<b>ESAIR</b> [Room: Confed 5] Third Workshop on Exploiting Semantic Annotations in Information Retrieval	<b>PIKM</b> [Room: Alberta] Workshop for Ph.D. Students in Information and Knowledge Management	<b>SMUC</b> [Room: Manitoba] 2nd International Workshop on Search and Mining User-generated Contents	
Afternoon	<b>CloudDB</b> [Room: Confed 3] Second International Workshop on Cloud Data Management	<b>DOLAP</b> [Room: Algonquin] ACM 13th International Workshop on Data Warehousing and OLAP	<b>ESAIR</b> [Room: Confed 5] Third Workshop on Exploiting Semantic Annotations in Information Retrieval	<b>PIKM</b> [Room: Alberta] Workshop for Ph.D. Students in Information and Knowledge Management	<b>SMUC</b> [Room: Manitoba] 2nd International Workshop on Search and Mining User-generated Contents	

# WorkShop Programs

## Fourth Workshop on Analytics for Noisy Unstructured Text Data (AND2010)

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

### Workshop Organizers:

1. Roberto Basili, University of Rome, Italy
2. Daniel Lopresti, Lehigh University, United States of America
3. Christoph Ringlstetter, University of Munich, Germany
4. Shourya Roy, Xerox India Innovation Hub, India
5. L. Venkata Subramaniam, IBM Research, India
6. 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, Luis Sarmiento, 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 Brusilovsky (University of Pittsburgh, USA)

**Program Schedule:**

9:00-10:00 Keynote

**The Metadata Challenge: Promoting Discovery, Access, and Usability for Online Books** (50 mins)

John Mark Ockerbloom (University of Pennsylvania)

Discussion (10 mins)

10:00-10:30 Coffee break

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

**The sBook: towards Social and Personalized Learning Experiences** (15 mins)

Myriam Ribière, Jérôme Picault and Sylvain Squedin (Alcatel-Lucent Bell Labs)

**Real-Time Document Collaboration Using iPads** (15 mins)

Jennifer Pearson (Swansea University) and George Buchanan (City University)

**HCI Design Principles for eReaders** (25 mins)

Jennifer Pearson (Swansea University), George Buchanan (City University) and Harold Thimbleby (Swansea University)

Discussion (10 mins)

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

**Ebooks Children Would Want to Read and Engage with** (15 mins)

Monica Landoni (University of Lugano)

**Working with First Nations: On-Demand Book Service** (15 mins)

Nadia Caidi and Margaret Lam (University of Toronto)

**Implementing New Knowledge Environments: Building Upon Research Foundations to Understand Books and Reading in the Digital Age** (15 mins)

Ray Siemens and Julie Meloni (University of Victoria)

Discussion (10 mins)

12:30-14:00 Lunch

- 14:00-15:00 Keynote  
**The Present and Future of Google Books** (50 mins)  
James Crawford (Google Books)  
  
Discussion (10 mins)
- 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 panelists and discussion  
Panelists:  
James Crawford (Google Books),  
John Mark Ockerbloom (University of Pennsylvania),  
other panelists - TBC

**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 (University 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-Chairs:

- 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 Bipartite 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 Ordonez (University of Houston, USA)

**Program Schedule:**



- 8:00-8:15 **Workshop welcome and introduction**
- 8:15-9:00 **Keynote Address**  
 Session Chair: C. Ordonez (U. 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 (U. Calabria)
- Using Ontologies to Discover Fact IDs** (23 mins)  
 Alberto Abello, Oscar Romero
- Multiversion Data Warehouse Constraints** (23 mins)  
 Ines Zouari, Faiza Ghazzi, Rafik Bouaziz
- A Methodology and Tool for Conceptual Designing a Data Warehouse from Ontology-based Sources** (16 mins)  
 Ladjel Bellatreche, Selma Khouri
- Graph Layout Readability in OLAP Diagrams for Conceptual Modeling of Data Warehouses** (16 mins)  
 Jesus Pardillo, Florian Mansmann.
- Towards the Definition of Spatial Data Warehouses Integrity Constraints with OCL** (16 mins)  
 Kamal Bouil, Sandro Bimonte, Hadj Mahboubi, Francois Pinet
- 10:30-11:00 Coffee Break
- 11:00-12:30 **Session 2: OLAP Query Processing**  
 Session Chair: A. Abello (U. Politecnica de Catalunya)
- Towards OLAP Query Reformulation in Peer-to-Peer Data Warehousing** (23 mins)  
 Matteo Golfarelli, Federica Mandreoli, Wilma Penzo, Stefano Rizzi, Elisa Turricchia
- Immediate Materialized Views with Outerjoins** (23 mins)  
 Anisoara Nica
- R-MESHJOIN for Near-Real-Time Data Warehousing** (23 mins)  
 Muhammad Asif Naeem, Gillian Dobbie, Gerald Weber, Shafiq Alam
- Repairing OLAP Queries in Databases with Referential Integrity Errors** (16 mins)  
 Javier Garcia-Garcia, Carlos Ordonez
- 12:30-14:00 Lunch
- 14:00-15:00 **Panel: Relational versus Non-Relational Database Systems for Data Warehousing**  
 Moderator: Carlos Ordonez (U. 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 (U. Libre de Bruxelles)
- Time-HOBI: Indexing Dimension Hierarchies by Means of Hierarchically Organized Bitmaps** (23 mins)  
 Jan Chmiel, Tadeusz Morzy, Robert Wrembel
- Exploring Graphics Processing Units as Parallel Coprocessors for Online Aggregation** (23 mins)  
 Tobias Lauer, Amitava Datta, Zurab Khadikov, Christoffer Anselm
- Integrating OLAP and Recommender Systems: An Evaluation Perspective** (23 mins)  
 Artus Krohn-Grimberghe, Alexandros Nanopoulos, Lars Schmidt-Thieme
- Balancing Accuracy and Privacy of OLAP Aggregations on Data Cubes** (16 mins)  
 Alfredo Cuzzocrea, Domenico Sacca
- Context-Aware Generalization for Cube Measures** (16 mins)  
 Yoann Pitarch, Cecile Favre, Anne Laurent, Pascal Poncelet
- 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:**

1. Jaap Kamps (University of Amsterdam)
2. Jussi Karlgren (SICS, Stockholm)
3. Ralf Schenkel (MPI/Saarland University)

**Program Schedule:**

- 9:00-9:30     **Welcome & 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 Text 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 Mladeníć 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)**

**Workshop 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  
**Speaker: 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 Klifman

**Adaptive Query Processing in Data Stream Management Systems under Limited Memory Resources**

Fatima Farag, Moustafa Hammad, Reda Alhajj

**Efficient Algorithms based on Relational Queries to Mine Frequent Graphs**

Walter Garcia, Carlos Ordonez and Kai Zhao

**Quality Factory and Quality Notification Service in Data Warehouse**

Yan L, 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
- A Decision Support System for Green Data Centers**  
Michael Pawlish, Aparna S. Varde
- Leveraging Natural Language Processing of Clinical Narratives for Phenotype Modeling**  
Preethi Raghavan, Albert Lai
- Entity Classification by Bag of Wikipedia Articles**  
Tomas Kliegr
- 15:30-16:00 Coffee Break and Poster Session
- 16:00-17:00 **Session 3: Information Retrieval Track (IR)**  
Session Chair: Anisoara Nica, Sybase iAnywhere, Waterloo, Canada
- Research Proposal for Distributed Deep Web Search**  
Kien Tjin-Kam-Jet
- From Exploratory Search to Web Search and Back**  
Roberto Mirizzi and Tommaso Di Noia
- 17:00-18:00 **Session 4: Poster Session**
- Identifying the Challenges for Optimizing the Process to Achieve Reproducibility of Results**  
Mohammad Rezwanaul 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 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 Cortizo (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, Luis Sarmiento 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, Egipt), Masoud Makrehchi (Thomson Reuters, USA), Jérôme Picault (Alcatel-Lucent Bell Labs), Haggai Rotman (IBM)