



# CIKM

NAPA VALLEY  
2008

ACM 17th Conference on Information and Knowledge Management

## Conference Program

### Editors

James G. Shanahan, Sihem Amer-Yahia, Yi Zhang  
Alek Kolcz, Abdur Chowdury, Diane Kelly



Sponsored by:

sig web

SIGIR

**General Chair:**

James G. Shanahan (*Church and Duncan Group Inc, USA*)

**Program Chairs (Database):**

Sihem Amer-Yahia (*Yahoo! Research, USA*)

Ioana Manolescu (*INRIA, France*)

**Program Chairs (Information Retrieval):**

Yi Zhang (*University of California, Santa Cruz, USA*)

David A. Evans (*JustSystems Evans Research, USA*)

**Program Chairs (Knowledge Management):**

Alek Kołcz (*Microsoft Live Labs, USA*)

Key-Sun Choi (*KAIST, Korea*)

**Program Chair (Industry):**

Abdur Chowdury (*Twitter, USA*)

**Proceedings Chair:**

Diane Kelly (*University of North Carolina at Chapel Hill, USA*)

**Tutorials Chair:**

Neel Sundaresan (*eBay Research Labs, USA*)

**Workshops Chair:**

Gregory Grefenstette (*Exalead, France*)

**Panels Chair:**

David A. Evans (*JustSystems Evans Research, USA*)

**Industry Day Chair:**

Marius Pasca (*Google, Inc., USA*)

**Local Area Chair:**

Goutham Kurra (*Turn, Inc, USA*)

**Sponsorship Chair:**

Yan Qu (*Platform-A, USA*)

**Treasurer:**

Ryen W. White (*Microsoft Research, USA*)

**Publicity Chair:**

Mounia Lalmas (*Queen Mary, University of London, UK*)

**Web Master:**

Disco Gerdes (*Twinnov Web Development, Argentina*)

# Organization

# Table of Contents

Foreword.....	3
Message from the Program Chairs.....	4
Sponsors.....	5
Program-at-a-glance.....	6
Tutorials.....	7
Keynotes.....	8
Sessions.....	12
Posters Rules & Best Poster.....	21
Posters.....	22
Industry Day.....	27
Panels.....	28
Workshops.....	29
Conference Venue Map.....	31
Internet Access.....	31
Banquet.....	32



# Foreword

Welcome to Napa Valley and to the 17th ACM Conference on Information and Knowledge Management—CIKM 2008.

CIKM has become the Cabernet Sauvignon of information science conferences; that is an audacious incongruous statement that I will try to shed some light on here and that will become more apparent at conference time. CIKM, like Cabernet Sauvignon the red wine grape, is a crossing between well established varieties (Cabernet franc and Sauvignon blanc); in this case Databases, Information Retrieval, Knowledge Management and Industry. Furthermore, echoing Cabernet Sauvignon, to make the CIKM 2008 varietal more palatable and to broaden its appeal, we have blended these traditional varieties with Industry Day and Panel Discussions just as Cabernet Sauvignon is often blended with Merlot, Cab. Franc, or Malbec.

The CIKM 2008 vintage, building on previous CIKMs, has become one of the world's most widely recognized varietals as supported by the following statistics:

- Almost 800 refereed paper submissions (previous years this has been about 500)
- Over 620 people have registered for the conference at the time of this writing (typically this has been around 300)
- 132 papers were accepted for presentation as full papers (17% acceptance rate) and an additional 103 as posters.

The Program Co-chairs Sihem Amer-Yahia, Yi Zhang, and Alek Kolcz are to be commended for assembling superb committees of Meta-Reviewers for each area (39 members) and paper reviewers (almost 500 in all), who worked very hard to provide excellent peer feedback to the authors in an on-time manner despite the 50% increase in submissions. This year we adapted, for the first time at CIKM, a two-tier peer review system; it worked really well. Thanks also to Abdur Chowdhury, who chaired the Industrial Track. I also thank Diane Kelly and David Evans who courageously stepped in when IR became ER in the wake of an unexpected event just days before our program committee meeting. Diane Kelly has done a great job in getting the proceedings together and has made the interaction with our publishers seamless.

I am very grateful to a fantastic roster of keynote speakers, invited speakers and panelists including Rakesh Agrawal, Bruce Croft and Pedro Domingos. Thanks in part to the Program Area Chairs, David Evans and Marius Pasca for making this happen.

We thank the authors for providing such a stimulating program of papers and posters.

Yan Qu has engaged a great team of sponsors and supporters who have taken the financial edge off of the conference organization;

needless to say without them this would have been impossible. Thanks to the following: sponsors: ACM, SIGIR and SIGWEB; Gold level supporters: Information Retrieval Facility, Microsoft, and Searchme; Silver level supporters: Platform-A, Ask.com, eBay Research Labs, and Yahoo! Research; Bronze level supporters: A9, JustSystems Evans Research, UMBC, Centre for Next Generation Localisation (CNGL), Exalead, LinkedIn and Turn; Book exhibits: Cambridge University Press and Springer; Video supporter: Videolectures.Net.

Organizing CIKM 2008 has been a fantastic team effort. The following deserve a special mention for the tireless energy and wealth of knowledge and ideas that they poured into the CIKM 2008 vintage: Goutham Kurra on Local Arrangements; Neel Sundaresan on tutorials; Greg Grefenstette on workshops; Mounia Lalmas on PR; Ryen W. White on finances; Disco Gerdes on website and conference program matters; and Erica E. Wray on all matters legal. Thanks to eBay, Microsoft and Yahoo! for hosting a variety of CIKM 2008 organizational meetings throughout the past 18 months.

CIKM 2008 owes a lot to the following who have been nothing short of fantastic in their time and guidance: Mario J. Silva, E.K. Park, Charles Nicholas, Ricardo Baeza-Yates, Andrei Broder, Francisco M. Couto, and Ophir Frieder.

This year we have granted 32 travel awards to students; Mounia Lalmas and Ryen W. White headed this effort. Student volunteers also deserve a word of appreciation for their assistance. I'd also like to thank Songphan Choemprayong (University of North Carolina), for designing and creating the conference program.

Incongruous in this old wine country setting, CIKM 2008 welcomes you, the hi-tech scientists from around the world, to exchange new ideas concerning searching, managing and exploring information. I wish every participant a very productive conference and hope that you also find time to enjoy the many attractions that Napa Valley (including its internationally renowned Cabernet Sauvignon) and its surrounding areas have to offer.

**James G. Shanahan**

CIKM 2008 General Chair  
*Church and Duncan Group, Inc.*

# Message from the Program Committee Chairs

It is our great pleasure to welcome you to the 17th ACM Conference on Information and Knowledge Management – CIKM'08. Since 1992, the ACM Conference on Information and Knowledge Management (CIKM) has been successfully bringing together leading researchers and developers from the database, information retrieval, and knowledge management communities. 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. In CIKM 2008, we continued the tradition of promoting collaboration among the general areas of databases, information retrieval, and knowledge management. This year's call for papers attracted almost 800 submissions from Asia, Canada, Europe, Africa, and the United States. The program committee accepted 132 papers and 103 posters giving CIKM'08 an acceptance rate of 17%.

Putting together CIKM 2008 was a team effort. First of all, we would like to thank the authors and panelists for providing the content of the program. We would like to express our gratitude to the program committee members as well as external reviewers, who worked very hard in reviewing papers and providing suggestions for improvement. We would also like to thank the countless others who have contributed to making CIKM'08 a success. Individual area assistant chairs were very helpful at various stages during the conference process. A special mention goes to Diane Kelly and David Evans without whom the final paper selection process would not have happened. We would also like to thank previous area chairs Alberto Laender, Ricardo Baeza-Yates and Deborah McGuinness for providing invaluable guidance and sometimes,

reviews! We would also like to thank the ACM and all our corporate sponsors for their generous support. Finally, we are all deeply thankful to James Shanahan, the General Chair, for his tremendous efforts in making this work a dynamic and enjoyable team effort. Thank you Jimi for your contagious energy!

We hope that you will find this program interesting and thought-provoking and that the symposium will provide you with a valuable opportunity to share ideas with other researchers and practitioners from around the world.

**Sihem Amer-Yahia**

CIKM 2008 DB Chair  
*Yahoo! Research, USA*

**Yi Zhang**

CIKM 2008 IR Chair  
*University of California,  
Santa Cruz, USA*

**Alek Koltz**

CIKM 2008 KM Chair  
*Microsoft Live Labs, USA*

**Abdur Chowdhury**

CIKM 2008 Industry  
Chair  
*Twitter, USA*

# Sponsors

## Gold Supporters



## Silver Supporters



## Bronze Supporters



## Video Supporters



## Book Exhibits



# Program-at-a-glance

	Sunday October 26	Monday October 27	Tuesday October 28	Wednesday October 29	Thursday October 30
8:00-8:30		Welcome and Opening Remarks <i>Grand Ballroom</i>			
8:30-9:30		Keynote <i>Grand Ballroom</i> Rakesh Agrawal <i>Humane Data Mining</i>	Keynote <i>Grand Ballroom</i> Pedro Domingos <i>Markov Logic: A Unifying Language for Information and Knowledge Management</i>	Keynote: <i>Grand Ballroom</i> W. Bruce Croft <i>Unsolved Problems in Search (and How We Might Approach Them)</i>	
9:30-9:45		Poster Boaster <i>Grand Ballroom</i> DB:Sihem Amer-Yahia IR:Yi Zhang KM:Alek Kolcz Industry:Abdur Chowdury	Poster Boaster <i>Grand Ballroom</i> DB:Sihem Amer-Yahia IR:Yi Zhang KM:Alek Kolcz	Poster Boaster <i>Grand Ballroom</i> DB:Sihem Amer-Yahia IR:Yi Zhang KM:Alek Kolcz	
9:45-10:15		Coffee Break & Posters			
10:15-11:45		1A: Faceted Search, Web Query Results Presentation <i>Syrah &amp; Merlot</i> 1B: IR: Web Search 1 <i>Cabernet &amp; Zinfandel</i> 1C: KM: Classification <i>Stags Leap &amp; Howell Mountain</i> 1D: Industry Research Track <i>Mt. Veeder &amp; Carneros</i>	4A: DB/Industry: XML Data Integration and XML Query Optimization <i>Mt. Veeder &amp; Carneros</i> 4B: IR: Evaluation <i>Syrah &amp; Merlot</i> 4C: KM: Statistical Techniques <i>Stags Leap &amp; Howell Mountain</i> 4D: Panel Discussion: E-Discovery <i>Cabernet &amp; Zinfandel</i>	7A: IR: Advertising & Filtering <i>Syrah &amp; Merlot</i> 7B: IR: Blog <i>Stags Leap &amp; Howell Mountain</i> 7C: KM: Clustering <i>Mt. Veeder &amp; Carneros</i> 7D: Industry Day (1) <i>Cabernet &amp; Zinfandel</i>	
11:45-13:30		Lunch & Posters			
13:30-15:00		2A: DB: Efficient Maintenance and Query Optimization <i>Syrah &amp; Merlot</i> 2B: IR: Social Search <i>Cabernet &amp; Zinfandel</i> 2C: IR/KM: Machine Learning <i>Stags Leap &amp; Howell Mountain</i> 2D: KM: Link and Graph Mining <i>Mt. Veeder</i> 2E: KM: Information Filtering <i>Carneros</i>	5A: DB: Indexing and Physical Query Optimization <i>Carneros</i> 5B: IR: Web Search 2 <i>Syrah &amp; Merlot</i> 5C: IR: Multilingual & Multimedia <i>Cabernet &amp; Zinfandel</i> 5D: KM: Data Mining <i>Stags Leap &amp; Howell Mountain</i> 5E: KM: Semantic Techniques <i>Mt. Veeder</i>	8A: IR: Enterprise Search <i>Syrah &amp; Merlot</i> 8B: IR: Structured Documents <i>Stags Leap &amp; Howell Mountain</i> 8C: KM: Text Mining <i>Mt. Veeder &amp; Carneros</i> 8D: Industry Day (2) <i>Cabernet &amp; Zinfandel</i>	
15:00-15:30		Coffee Break & Posters			
15:30-17:00		3A: DB: Stream Processing <i>Mt. Veeder &amp; Carneros</i> 3B: IR: Theory <i>Stags Leap &amp; Howell Mountain</i> 3C: IR: Query Analysis <i>Syrah &amp; Merlot</i> 3D: KM: Web Mining <i>Cabernet &amp; Zinfandel</i>	6A: DB: Security and Privacy <i>Mt. Veeder</i> 6B: IR: Medley <i>Carneros</i> 6C: IR: Recommender Systems <i>Syrah &amp; Merlot</i> 6D: KM: Feature Selection <i>Stags Leap &amp; Howell Mountain</i> 6E: Panel Discussion 2: The Social (Open) Workspace <i>Cabernet &amp; Zinfandel</i>	9A: DB: Mobile and Distributed Data Management <i>Mt. Veeder &amp; Carneros</i> 9B: IR: Question Answering <i>Stags Leap &amp; Howell Mountain</i> 9C: KM: Information Extraction <i>Syrah &amp; Merlot</i> 9D: Industry Day (3) <i>Cabernet &amp; Zinfandel</i>	
		18:00 - 21:00 Searchme: Opening Reception <i>Pool Courtyard</i> Spanish party	17:45 - 22:30 Banquet <i>COPIA</i> 500 1st. St. <i>Napa, CA 94559</i>	17:15 - 18:00 Townhall Meeting <i>Grand Ballroom</i> 18:00 - 19:00 Wine reception <i>Corridor &amp; Pool Courtyard</i>	

Tutorials — See page 7 for schedule

Workshops — See pages 29 - 30 for schedule

# Tutorials

**Sunday, October 26**

## Full Day (11:00-19:00)

12:30-13:30: Lunch Break (not provided)  
13:00: Coffee, tea, juices, and pastries  
17:00-17:30: Coffee Break

### Large Graph Mining: Patterns, Tools and Case Studies

Location: *Howell Mountain*

Christos Faloutsos, Hanghang Tong (*Carnegie Mellon University, USA*)

### Log-linear Models and Conditional Random Fields

Location: *Atlas Peak*

Charles Elkan (*University of California, San Diego, USA*)

## Morning (11:00-15:00)

12:30-13:30: Lunch Break (not provided)  
13:00: Coffee, tea, juices, and pastries

### Evolution of Rule-based Information Extraction: from Grammars to Algebra

Location: *Mt. Veeder*

Rajasekar Krishnamurthy, Sriram Raghavan, Huaiyu Zhu (*IBM Almaden Research Center, USA*)

### Information and Knowledge Management with Graphs and Matrices

Location: *Diamond Mountain*

Fei Wang (*Tsinghua University, China*), Tao Li (*Florida International University, USA*), Chris Ding (*University of Texas at Arlington, USA*)

### Machine Learning for Information Retrieval

Location: *Castle Peak*

Rong Jin (*Michigan State University, USA*), Yi Zhang (*University of California, Santa Cruz, USA*)

## Afternoon (15:30-19:00)

17:00-17:30: Coffee Break

### Web Search Log Analysis and User Behavior Modeling

Location: *Diamond Mountain*

Peiling Wang, Lei Wu (*University of Tennessee, USA*), Dietmar Wolfram (*University of Wisconsin-Milwaukee, USA*)

### Algorithmic Challenges in Online Advertising

Location: *Mt. Veeder*

Deepak Agarwal, Deepayan Chakrabarti (*Yahoo! Research, USA*)



# *Keynotes*



**Rakesh Agrawal**  
Microsoft Research, USA

## Humane Data Mining

Monday, October 27

8:30-9:30

Location: Grand Ballroom

Data Mining has made tremendous strides in the last decade. It is time to take data mining to the next level of contributions, while continuing to innovate for the current mainstream market. We postulate that a fruitful future direction could be humane data mining: applications to benefit individuals. The potential applications include personal data mining (e.g. personal health), enable people to get a grip on their world (e.g., dealing with the long tail of search), enable people to become creative (e.g., inventions arising from linking non-interacting scientific literature), enable people to make contributions to society (e.g. education collaboration networks), and data-driven science (e.g., study ecological disasters, brain disorders). Rooting our future work in these (and similar) applications will lead to new data mining abstractions, algorithms, and systems.

Rakesh Agrawal is a Microsoft Technical Fellow and heads the Search Labs in Microsoft Research. He is the recipient of the ACM-SIGKDD First Innovation Award, ACM-SIGMOD Edgar F. Codd Innovations Award, ACM-SIGMOD Test of Time Award, VLDB 10-Yr Most Influential Paper Award, ICDE Most Influential Paper Award, and Computerworld First Horizon Award. He is a Member of the National Academy of Engineering, a Fellow of ACM, and a Fellow of IEEE. Scientific American named him to the list of 50 top scientists and technologists in 2003.

Prior to joining Microsoft in March 2006, Rakesh was an IBM Fellow and led the Quest group at the IBM Almaden Research Center. Earlier, he was with the Bell Laboratories, Murray Hill from 1983 to 1989. He also worked for 3 years at India's premier company, the Bharat Heavy Electricals Ltd. He received the M.S. and Ph.D. degrees in Computer Science from the University of Wisconsin-Madison in 1983. He also holds a B.E. degree in Electronics and Communication Engineering from IIT-Roorkee, and a two-year Post Graduate Diploma in Industrial Engineering from the National Institute of Industrial Engineering (NITIE), Bombay.

Rakesh is well-known for developing fundamental data mining concepts and technologies and pioneering key concepts in data privacy, including Hippocratic Database, Sovereign Information Sharing, and Privacy-Preserving Data Mining. IBM's commercial data mining product, Intelligent Miner, grew out of his work. His research has been incorporated into other IBM products, including DB2 Mining Extender, DB2 OLAP Server and WebSphere Commerce Server, and has influenced several other commercial and academic products, prototypes and applications. His other technical contributions include Polyglot object-oriented type system, Alert active database system, Ode (Object database and environment), Alpha (extension of relational databases with generalized transitive closure), Nest distributed system, transaction management, and database machines.

Rakesh has been granted 60 patents. He has published more than 150 research papers, many of them considered seminal. He has written the 1st as well as 2nd highest cited of all papers in the fields of databases and data mining (13th and 15th most cited across all computer science as of February 2007 in CiteSeer). Wikipedia lists one

of his papers as one of the most influential database papers. His papers have been cited more than 6500 times, with more than 15 of them receiving more than 100 citations each. He is the most cited author in the field of database systems. His work has been featured in New York Times Year in Review, New York Times Science section, and several other publications.

## Markov Logic: A Unifying Language for Information and Knowledge Management

Tuesday, October 28

8:30-9:30

Location: Grand Ballroom

**Pedro Domingos**

Department of Computer Science and Engineering  
University of Washington



Modern information and knowledge management is characterized by high degrees of complexity and uncertainty. Complexity is well handled by first-order logic, and uncertainty by probabilistic graphical models. What has been sorely missing is a seamless combination of the two. Markov logic provides this by attaching weights to logical formulas and treating them as templates for features of Markov random fields. I will survey Markov logic representation, inference, learning and applications. Inference algorithms combine ideas from satisfiability testing, resolution, Markov chain Monte Carlo and belief propagation. Learning algorithms involve statistical weight learning and inductive logic programming. Markov logic has been successfully applied to a wide range of information and knowledge management problems, including information extraction, entity resolution, ontology learning, link prediction, heterogeneous knowledge bases, and others. It is the basis of the open-source Alchemy system. (Joint work with Stanley Kok, Daniel Lowd, Hoifung Poon, Matt Richardson, Parag Singla, Marc Sumner, and Jue Wang.)

Pedro Domingos is Associate Professor of Computer Science and Engineering at the University of Washington. His research interests are in artificial intelligence, machine learning and data mining. He received a PhD in Information and Computer Science from the University of California at Irvine, and is the author or co-author of over 100 technical publications. He

is a member of the advisory board of JAIR, a member of the editorial board of the Machine Learning journal, and a co-founder of the International Machine Learning Society. He was program co-chair of KDD-2003, and has served on numerous program committees. He has received several awards, including a Sloan Fellowship, an NSF CAREER Award, a Fulbright Scholarship,

an IBM Faculty Award, and best paper awards at KDD-98, KDD-99 and PKDD-2005.



**W. Bruce Croft**

Department of Computer Science,  
University of Massachusetts Amherst

## Unsolved Problems in Search (and how we might approach them)

Wednesday, October 29

8:30-9:30

Location: Grand Ballroom

Search applications have become ubiquitous and very successful. Major advances have been made in understanding how to deliver effective results very efficiently for a class of queries. As the range of applications broaden to include Web search, desktop search, enterprise search, vertical search, social search, etc., the number of new research challenges has appeared to grow rather than shrink. Many of these challenges are variations on underlying themes and principles that information retrieval has focused on for more than 40 years. In this talk, the unsolved problems arising from new search applications will be categorized and discussed in terms of information retrieval models, and some potential paths to solutions for these problems will be outlined.

W. Bruce Croft is a Distinguished Professor in the Department of Computer Science at the University of Massachusetts, Amherst, which he joined in 1979. In 1992, he founded the Center for Intelligent Information Retrieval (CIIR), which combines basic research with technology transfer to a variety of government and industry partners. Dr. Croft was Chair of the department from 2001-2007.

He received the B.Sc.(Honors) degree in 1973, and an M.Sc. in Computer Science in 1974 from Monash University in Melbourne, Australia. His Ph.D. in Computer Science was from the University of Cambridge, England in 1979.

His research interests are in many areas of information retrieval, including retrieval models, representation, Web search, query processing, cross-lingual retrieval, and search architectures. He has published more than 200 articles on these and other subjects, has served on numerous program committees, and has been involved in the organization of many workshops and conferences.

Dr. Croft was a member of the National Research Council Computer Science and Telecommunications Board, 2000-2003, and Editor-in-Chief of ACM Transactions on Information Systems, 1995-2002. Dr. Croft was

elected a Fellow of ACM in 1997, received the Research Award from the American Society for Information Science and Technology in 2000, and received the Gerard Salton Award from the ACM Special Interest Group in Information Retrieval (SIGIR) in 2003.

**Poster Boaster** (Session Chair: Sihem Amer-Yahia)

*Grand Ballroom*

DB: Sihem Amer-Yahia

IR: Yi Zhang

KM: Alek Kolcz

Industry: Abdur Chowdury

**SESSION 1**

Monday, October 27 (10:15-11:45)

**1A - DB: Faceted Search, Web Query Results Presentation** (Session Chair: Jimeng Sun)

*Location: Syrah & Merlot*
**Dynamic Faceted Search for Discovery-Driven Analysis** Dash Debabrata, Rao Jun, Nimrod Megiddo, Anastasia Ailamaki, Guy Lohman (*Carnegie Mellon University and IBM Almaden Research Center, USA*)

**Minimum Effort Driven Dynamic Faceted Search in Structured Databases** Senjuti Basu Roy, Haidong Wang, Gautam Das, Ullas Nambiar, Mukesh Mohania (*University of Texas at Arlington, USA*)

**A Language for Manipulating Clustered Web Documents Results** Gloria Bordogna, Alessandro Campi, Giuseppe Psaila, Stefania Ronchi (*CNR IDPA and University of Bergamo, Italy*)

**Integrating Web Query Results: Holistic Schema Matching** Shui-Lung Chuang, Kevin Chang (*University of Illinois at Urbana-Champaign, USA*)

**1B - IR: Web Search 1** (Session Chair: Susan Dumais)

*Location: Cabernet & Zinfandel*
**How does Clickthrough Data Reflect Retrieval Quality?** Filip Radlinski, Madhu Kurup, Thorsten Joachims (*Cornell University, USA*)

**Efficient and Effective Link Analysis with Precomputed SALSA Maps** Marc Najork, Nick Craswell (*Microsoft Research Cambridge, UK*)

**Achieving both High Precision and High Recall in Near-Duplicate Detection** Lianen Huang, Lei Wang, Xiaoming Li (*Peking University, China*)

**Are Clickthrough Data Adequate for Learning Web Search Rankings?** Zhicheng Dou, Ruihua Song, Xiaojie Yuan, Ji-Rong Wen (*Nankai University, Shanghai Jiao Tong University and Microsoft Research Asia, China*)

**1C - KM: Classification** (Session Chair: Iadh Ounis)

*Location: Stags Leap & Howell Mountain*
**Error-Driven Generalist&Experts (EDGE): A Multi-stage Ensemble Framework for Text Categorization** Jian Huang, Omid Madani, C. Lee Giles (*Pennsylvania State University, USA*)

**A Sparse Gaussian Processes Classification Framework for Fast Tag Suggestions** Yang Song, Lu Zhang, C. Lee Giles (*Pennsylvania State University, USA*)

**Transfer Learning From Multiple Source Domains via Consensus Regularization** Ping Luo, Fuzhen Zhuang, Hui Xiong, Yuhong Xiong, Qing He (*Chinese Academy of Science, China*)

**Classifying Networked Entities with Modularity Kernels** Dell Zhang, Robert Mao (*University of London, UK*)

**1D - Industry Research Track** (Session Chair: Nazli Goharian)

*Location: Mt. Veeder & Carneros*
**Web-Scale Named Entity Recognition** Casey Whitelaw, Alex Kehlenbeck, Nemanja Petrovic, Lyle Ungar (*University of Pennsylvania, USA*)

**Semi-Automated Logging of Contact Center Telephone Calls** Roy Byrd, Mary Neff, Wilfried Teiken, Youngja Park, Keh-Shin F Cheng, Stephen Gates, Karthik Viswesvariah (*IBM T.J. Watson Research Center, USA*)

**MedSearch: A Specialized Search Engine for Medical Information** Gang Luo, Chunqiang Tang, Hao Yang, Xin Wei (*IBM T.J. Watson Research Center, USA*)

**An Empirical Study of Required Dimensionality for Large-scale Latent Semantic Indexing Applications** Roger Bradford (*Agilex Technologies Inc., USA*)

**2A - DB: Efficient Maintenance and Query Optimization** (Session Chair: Luke Huan)Location: *Syrac & Merlot***Content-Based Filtering for Efficient Online Materialized View Maintenance** Gang Luo, Philip Yu (*IBM T.J. Watson Research Center, USA*)**A Step towards Incremental Maintenance of Composed Schema Mappings** Qian Gang (*Nanjing University of Finance & Economics, China*)**Modeling and Exploiting Query Interactions in Database Systems** Mumtaz Ahmad, Ashraf Aboulmaga, Shivnath Babu, Kamesh Munagala (*University of Waterloo, Canada and Duke University, USA*)**A Novel Optimization Approach to Efficiently Process Aggregate Similarity Queries in Metric Access Methods** Humberto Razente, Maria Camila Barioni, Agma Traina, Christos Faloutsos, Caetano Traina (*University of São Paulo at São Carlos, Brazil*)**2B - IR: Social Search** (Session Chair: Neel Sundaresan)Location: *Cabernet & Zinfandel***Can All Tags Be Used for Search?** Kerstin Bischoff, Claudiu Firan, Wolfgang Nejdl, Raluca Paiu (*University of Hannover, Germany*)**Comparing Citation Contexts for Information Retrieval** Anna Ritchie, Simone Teufel, Stephen Robertson (*University of Cambridge, UK*)**Social Tags: Meanings and Suggestions** Fabian Suchanek, Milan Vojnovic, Dinan Gunawardena (*Max-Planck-Institute for Computer Science, Germany*)**Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection** Hao Ma, Haixuan Yang, Michael R. Lyu, Irwin King (*The Chinese University of Hong Kong, China*)**2C - IR/KM: Machine Learning** (Session Chair: Rong Jin)Location: *Stags Leap & Howell Mountain***Exploiting Temporal Contexts in Text Classification** Leonardo Rocha, Fernando Mourão, Adriano Pereira, Marcos Gonçalves, Wagner Meira (*Federal University of Minas Gerais, Brazil*)**Kernel Methods, Syntax and Semantics for Relational Text Categorization** Alessandro Moschitti (*University of Trento, Italy*)**BNS Feature Scaling: An Improved Representation over TF-IDF for SVM Text Classification** George Forman (*HP Labs, USA*)**Learning a Two-Stage SVM/CRF Sequence Classifier** Guilherme Hoefel, Charles Elkan (*University of California San Diego, USA*)**2D - KM: Link and Graph Mining** (Session Chair: Christos Faloutsos)Location: *Mt. Veeder***Local Approximation of PageRank and Reverse PageRank** Li-Tal Mashiach, Ziv Bar-Yossef (*Technion, Israel*)**Link Privacy in Social Networks** Aleksandra Korolova, Rajeev Motwani, Shubha U. Nabar, Ying Xu (*Stanford University, USA*)**On Effective Presentation of Graph Patterns: A Structural Representative Approach** Chen Chen, Cindy Lin, Xifeng Yan, Jiawei Han (*University of Illinois at Urbana-Champaign, USA*)**Characterizing and Predicting Community Members from Evolutionary and Heterogeneous Networks** Sourav S. Bhowmick, Qiankun Zhao, Xin Zheng, Kai Yi (*Nanyang Technological University, Singapore*)**2E - KM: Information Filtering** (Session Chair: Lyle Ungar)Location: *Carneros***An Algorithm to Determine Peer-Reviewers** Marko Rodriguez, Johan Bollen (*Los Alamos National Laboratory, USA*)**Spam Characterization and Detection in Peer-to-Peer File-Sharing Systems** Dongmei Jia, Wai Gen Yee, Ophir Frieder (*Illinois Institute of Technology, USA*)**Predicting Web Spam with HTTP Session Information** Steve Webb, James Caverlee, Calton Pu (*Georgia Institute of Technology and Texas A&M University, USA*)**Inferring Semantic Query Relations from Collective User Behavior** Nish Parikh, Neel Sundaresan (*eBay Research Labs, USA*)

**3A-DB: Stream Processing** (Session Chair: Ashraf Aboulnaga)Location: *Mt. Veeder & Carneros***Anomaly-Free Incremental Output in Stream Processing** George Mihaila, Ioana Roxana Stanoi, Christian Lang (*IBM T.J. Watson Research Center and IBM Almaden Research Center, USA*)**SNIF TOOL: Sniffing for Patterns in Continuous Streams** Abhishek Mukherji, Elke Rundensteiner, David Brown, Venkatesh Raghavan (*Worcester Polytechnic Institute, USA*)**Real-Time New Event Detection for Video Streams**Gang Luo, Rong Yan, Philip Yu (*IBM T.J. Watson Research Center, USA*)**Linear Time Membership in a Class of Regular Expressions with Interleaving and Counting** Giorgio Ghelli, Dario Colazzo, Carlo Sartiani (*Università di Pisa, Italy*)**3B - IR: Theory** (Session Chair: Michael Lyu)Location: *Stags Leap & Howell Mountain***Generalized Inverse Document Frequency** Donald Metzler (*Yahoo! Research, USA*)**TinyLex: Static N-Gram Index Pruning with Perfect Recall** Derrick Coetzee (*Microsoft Research, USA*)**Revisiting the Relationship between Document Length and Relevance** David Losada, Leif Azzopardi, Mark Baillie (*University of Santiago de Compostela, Spain, University of Glasgow and University of Strathclyde, UK*)**Relating Dependent Indexes using Dempster-Shafer Theory** Lixin Shi, Jian-Yun Nie, Guihong Cao (*University of Montreal, Canada*)**3C - IR: Query Analysis** (Session Chair: Diane Kelly)Location: *Syrah & Merlot***Improved Query Difficulty Prediction for the Web**Claudia Hauff, Vanessa Murdock, Ricardo Baeza-Yates (*University of Twente, The Netherlands and Yahoo! Research Barcelona, Spain*)**Understanding the Relationship between Searchers' Queries and Information Goals** Doug Downey, Dan Liebling, Susan Dumais (*University of Washington and Microsoft Research, USA*)**Active Relevance Feedback for Difficult Queries** Zuobing Xu, Ram Akella (*University of California Santa Cruz, USA*)**Query Suggestion Using Hitting Time** Qiaozhu Mei, Dengyong Zhou, Kenneth Church (*University of Illinois at Urbana-Champaign, USA*)**3D - KM: Web Mining** (Session Chair: Jiang-Ming Yang)Location: *Cabernet & Zinfandel***Mining Term Association Patterns from Search Logs for Effective Query Reformulation** Xuanhui Wang, ChengXiang Zhai (*University of Illinois at Urbana-Champaign, USA*)**Non-Local Evidence for Expert Finding** Krisztian Balog, Maarten de Rijke (*University of Amsterdam, The Netherlands*)**Discovering Leaders from Community Actions** Amit Goyal, Francesco Bonchi, Laks V. S. Lakshmanan (*University of British Columbia, Canada and Yahoo! Research Barcelona, Spain*)**Learning to Link with Wikipedia** David Milne, Ian H. Witten (*University of Waikato, New Zealand*)

**Poster Boaster** (Session Chair: Alek Kolcz)

Grand Ballroom

DB: Sihem Amer-Yahia

IR: Yi Zhang

KM: Alek Kolcz

**SESSION 4**

Tuesday, October 28 (10:15-11:45)

**4A - DB/Industry: XML Data Integration and XML Query Optimization** (Session Chair: Caetano Traina)Location: *Mt. Veeder & Carneros***Rewriting of Visibly Pushdown Languages for XML Data Integration** Alex Thomo, Venkatesh Srinivasan (*University of Victoria, Canada*)**Some Rewrite Optimizations of XQuery Navigation in DB2** Jarek Gryz, Guangjun Xie, Qi Cheng, Calisto Zuzarte (*York University, Canada*)**Pruning Nested XQuery Queries** Billel Gueni, Talel Abdessalem, Bogdan Cautis, Emmanuel Waller (*LTCI - Telecom ParisTech and Université de Paris-Sud, France*)**Heuristic Approaches for Checking Containment of Generalized Tree-Pattern Queries** Pawel Placek, Dimitri Theodoratos, Stefanos Soudatos, Theodore Dalamagas, Timos Sellis (*New Jersey Institute of Technology, USA*)**4B - IR: Evaluation** (Session Chair: Ian Soboroff)Location: *Syrah & Merlot***Retrievability: An Evaluation Measure for Higher Order Information Access Tasks** Leif Azzopardi, Vishwa Vinay (*University of Glasgow, UK and Microsoft Research Labs Cambridge, UK*)**Statistical Power in Retrieval Experimentation** William Webber, Alistair Moffat, Justin Zobel (*University of Melbourne, Australia*)**Comparing Metrics across TREC and NTCIR: The Robustness to System Bias** Sakai Tetsuya (*NewsWatch, Inc., Japan*)**How Evaluator Domain Expertise Affects Search Result Relevance Judgements** Kenneth Kinney, Scott Huffiman, Juting Zhai (*Google, Inc., USA*)**4C - KM: Statistical Techniques** (Session Chair: Omid Madani)Location: *Stags Leap & Howell Mountain***Clustered Subset Selection and its Applications on IT Service Metrics** Christos Boutsidis, Jimeng Sun, Nikos Anerousis (*IBM T.J. Watson Research and Carnegie Mellon University, USA*)**The Query-Flow Graph: Model and Applications** Paolo Boldi, Francesco Bonchi, Carlos Castillo, Debora Donato, Aristides Gionis, Sebastiano Vigna (*Yahoo! Research Barcelona, Spain*)**A Framework for Estimating Complex Probability Density Structures in Data Stream** Arnold Boedihardjo, Chang-Tien Lu, Chen Feng (*Virginia Tech, USA*)**Proactive Learning: Cost-Sensitive Active Learning with Multiple Imperfect Oracles** Pinar Donmez, Jaime Carbonell (*Carnegie Mellon University, USA*)**4D - Panel Discussion: E-Discovery** (Session Chair: David A. Evans)Location: *Cabernet & Zinfandel***Why E-Discovery is a CIKM-Hard Problem** David A. Evans, President, CEO, and Chief Scientist (*JustSystems Evans Research, USA*)**Panning for Gold in E-Discovery: What Every Information Scientist Should Know About the Way Lawyers Search for Electronic Evidence** Jason R. Baron, Director of Litigation (*National Archives and Records Administration, USA*)**IR Perspectives on the E-Discovery Problem** Chris Buckley, President (*Sabir Research, USA*)**Technical Cases Studies from the E-Discovery Front Lines** Robert S. Bauer, Chief Technology Officer (*H5, USA*)



**5A - DB: Indexing and Physical Query Optimization** (Session Chair: Agma Traina)Location: *Carneros***Exploiting Pipeline Interruptions for Efficient Memory Allocation** Joseph Aguilar Saborti, Mohammad Jalali, Dave Sharpe, Victor Munteș-Mulero (*IBM, USA* and *Universitat Politècnica de Catalunya, Spain*)**A New Method for Indexing Genomes Using On-Disk Suffix Trees** Marina Barsky, Ulrike Stege, Alex Thomo, Chris Upton (*University of Victoria, Canada*)**Supporting Sub-Document Updates and Queries in an Inverted Index** Vuk Ercegova, Vanja Josifovski, Ning Li, Mauricio Mediano, Eugene Shekita (*IBM Almaden Research Center* and *Yahoo! Research, USA*)**Modeling LSH for Performance Tuning** Wei Dong, Zhe Wang, William Josephson, Moses Charikar, Kai Li (*Princeton University, USA*)**5B - IR: Web Search 2** (Session Chair: Thorsten Joachims)Location: *Syrac & Merlot***Can Phrase Indexing Help to Process Non-Phrase Queries?** Mingjie Zhu, Shuming Shi (*University of Science and Technology of China* and *Microsoft Research Asia, China*)**Matching Task Profiles and User Needs in Personalized Web Search** Julia Luxenburger, Shady Elbassuoni, Gerhard Weikum (*Max-Planck Institute for Informatics, Germany*)**Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs**Rosie Jones, Kristina Klinkner (*Yahoo! Research, USA*)**Learning Latent Semantic Relations from Clickthrough Data for Query Suggestion** Hao Ma, Haixuan Yang, Irwin King, Michael R. Lyu (*The Chinese University of Hong Kong, China*)**5C - IR: Multilingual & Multimedia** (Session Chair: Jian-Yun Nie)Location: *Cabernet & Zinfandel***Simultaneous Multilingual Search for Translingual Information Retrieval** Kristen Parton, Kathleen McKeown, James Allan, Enrique Henes-troza (*Columbia University* and *University of Massachusetts Amherst, USA*)**Translation Enhancement: A New Relevance Feedback Method for Cross-Language Information Retrieval** Daqing He, Dan Wu (*University of Pittsburgh, USA*)**High-Dimensional Descriptor Indexing for Large Multimedia Databases** Eduardo Valle, Matthieu Cord, Sylvie Philipp-Foliguet (*ETIS, France*)**On Low Dimensional Random Projections and Similarity Search** Yu-En Lu, Pietro Lio, Steven Hand (*University of Cambridge, UK*)**5D - KM: Data Mining** (Session Chair: Raj Bhatnagar)Location: *Stags Leap & Howell Mountain***Fast Mining of Complex Time-Stamped Events** Hanghang Tong, Yasushi Sakurai, Tina Eliassi-Rad, Christos Faloutsos (*Carnegie Mellon University, USA*)**Predicting Individual Disease Risk Based on Medical History** Darcy Davis, Nitesh Chawla, Nicholas Christakis, Nicholas Blumm, Laszlo Barabasi (*University of Notre Dame, USA*)**Identification of Gene Function Using Prediction by Partial Matching (PPM) Language Models** Malika Mahoui, W. Tehan, Arvind Kumar Thirumalaiswamy Sekhar, S. Chilukuri (*IUPUI, Bangor University, Dow AgroSciences, India*)**Fast Correlation Analysis on Time Series Datasets** Philon Nguyen, Nematollah Shiri (*Concordia University, USA*)**5E - KM: Semantic Techniques** (Session Chair: Marko Rodriguez)Location: *Mt. Veeder***Wildcards for Lightweight Information Integration in Virtual Desktops** Rodolfo Stecher, Claudia Niederee, Wolfgang Nejdl (*University of Hannover, Germany*)**Finding Informative Commonalities in Concept Collections** Simona Colucci, Eugenio Di Sciascio, Francesco Donini, Eufemia Tinelli (*D.O.O.M. s.r.l., Italy*)**Association Thesaurus Construction Methods based on Link Co-occurrence Analysis for Wikipedia** Masahiro Ito, KotaroNakayama, Takahiro Hara, Shojiro Nishio (*Osaka University, Japan*)**Peer Production of Structured Knowledge - An Empirical Study of Ratings and Incentive Mechanisms** Christian Huetter, Conny Kuehne (*University of Karlsruhe, Germany*)

**6A - DB: Security and Privacy** (Session Chair: Ioana Stanoi)Location: *Mt. Veeder***Efficient Techniques for Document Sanitization** Venkatesan Chakaravarthy, Himanshu Gupta, Prasan Roy, Mukesh Mohania (*IBM India Research Lab, India*)**Vanity Fair: Privacy in Querylog Bundles** Rosie Jones, Ravi Kumar, Bo Pang, Andrew Tomkins (*Yahoo! Research, USA*)**Dual Encryption for Query Integrity Assurance** Haixun Wang, Jian Yin, Chang-Shing Perng, Philip Yu (*IBMT.J.Watson Research Center, USA*)**Records Retention in Relational Database Systems** Ahmed Ataullah, Frank Tompa, Ashraf Aboulnaga (*University of Waterloo, Canada*)**6B - IR: Medley** (Session Chair: Jimmy Huang)Location: *Carneros***Joke Retrieval: Recognizing the same Joke told Differently.** Lisa Friedland, James Allan (*University of Massachusetts Amherst, USA*)**Ranked Feature Fusion Models for Ad Hoc Retrieval** Jeremy Pickens, Gene Golovchinsky (*FX Palo Alto Lab, Inc., USA*)**AdaSum: An Adaptive Model for Summarization** Zhang Jin (*Chinese Academy of Sciences, China*)**Modeling Hidden Topics on Document Manifold** Cai Deng (*University of Illinois at Urbana Champaign, USA*)**6C - IR: Recommender Systems** (Session Chair: Irwin King)Location: *Syrac & Merlot***Tapping on the Potential of Q&A Community by Recommending Answer Providers** Jinwen Guo, Shengliang Xu, Shenghua Bao, Yong Yu (*Shanghai Jiao Tong University, China*)**SoRec: Social Recommendation Using Probabilistic Matrix Factorization** Hao Ma, Haixuan Yang, Michael R. Lyu, Irwin King (*Chinese University Hong Kong, China*)**Probabilistic Polyadic Factorization and Its Application to Personalized Recommendation** Yun Chi, Shenghuo Zhu, Yihong Gong, Yi Zhang (*NEC Laboratories America and University of California Santa Cruz, USA*)**A Random Walk on the Red Carpet: Rating Movies with User Reviews and PageRank** Derry Wijaya, Stephane Bressan (*National University of Singapore, Singapore*)**6D - KM: Feature Selection** (Session Chair: Monica Rogati)Location: *Stags Leap & Howell Mountain***REDUS: Finding Reducible Subspaces in High Dimensional Data** Xiang Zhang, Feng Pan, Wei Wang (*University of North Carolina at Chapel Hill, USA*)**Mining Influential Attributes that Capture Class and Group Contrast Behaviour** Elsa Loekito, James Bailey (*University of Melbourne, Australia*)**Real-Time Data Pre-Processing Technique for Efficient Feature Extraction in Large Scale Data** Ying Liu (*Pennsylvania State University, USA*)**Structure Feature Selection for Graph Classification** Hongliang Fei, Jun Huan (*University of Kansas, USA*)**6E - Panel Discussion 2: The Social (Open) Workspace** (Session Chair: David A. Evans)Location: *Cabernet & Zinfandel***Automating Knowledge** Susan Feldman, Research Vice President (*Search and Digital Marketplace Technologies, IDC, USA*)**How to Augment Social Cognition** Ed H. Chi, Area Manager and Senior Research Scientist (*Palo Alto Research Center (PARC), USA*)**Trust and Design in Social Systems** Natasa Milic-Frayling, Principal Researcher and Director of Research Partnership Programme (*Microsoft Research Cambridge, UK*)**Using Social Networks for Social Work** Igor Perisic, Director of Search (*LinkedIn, USA*)

**Poster Boaster** (Session Chair: Yi Zhang)

*Grand Ballroom*

DB: Sihem Amer-Yahia

IR: Yi Zhang

KM: Alek Kolcz

**SESSION 7**

Wednesday, October 29 (10:15-11:45)

**7A - IR: Advertising & Filtering** (Session Chair: Tetsuya Sakai)

*Location: Syrah & Merlot*

**To Swing or not to Swing: Learning When (not) to Advertise** Andrei Broder, Massimiliano Ciaramita, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Donald Metzler, Vanessa Murdock, Vassilis Plachouras (*Yahoo! Research, USA, Pontificia Universidade Católica do Rio de Janeiro, Brazil and Yahoo! Research Barcelona, Spain*)

**Search Advertising using Web Relevance Feedback** Andrei Broder, Peter Ciccolo, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Lance Riedel (*Yahoo! Research, USA and Pontificia Universidade Católica do Rio de Janeiro, Brazil*)

**A Two-stage Text Mining Model for Information Filtering** Yuefeng Li, Xujuan Zhou, Peter Bruza, Yue Xu, Raymond Y.K. Lau (*Queensland University of Technology, Australia*)

**Automatic Online News Topic Ranking Using Media Focus and User Attention Based on Aging Theory** Canhui Wang, Min Zhang, Liyun Ru, Shaoping Ma (*Tsinghua University, China*)

**7B - IR: Blog** (Session Chair: Belle Tseng)

*Location: Stags Leap & Howell Mountain*

**Key Blog Distillation: Ranking Aggregates** Craig Macdonald, Iadh Ounis (*University of Glasgow, UK*)

**Blog Site Search Using Resource Selection** Jangwon Seo, Bruce Croft (*University of Massachusetts Amherst, USA*)

**An Effective Statistical Approach to Blog Post Opinion Retrieval** Ben He, Craig Macdonald, Jiyin He, Iadh Ounis (*University of Glasgow, UK and University of Amsterdam, NL*)

**7C - KM: Clustering** (Session Chair: Nitesh Chawla)

*Location: Mt. Veeder & Carneros*

**A Consensus Based Approach to Constrained Clustering of Software Requirements** Chuan Duan, Jane Cleland-Huang and Bamshad Mobasher (*DePaul University, USA*)

**Data Weaving: Scaling Up the State of the Art in Data Clustering** Ron Bekkerman, Martin Scholz (*HP Labs, USA*)

**EDSC: Efficient Density-Based Subspace Clustering** Ira Assent, Ralph Krieger, Emmanuel Müller, Thomas Seidl (*RWTH Aachen University, Germany*)

**An Effective Algorithm for Mining 3-Clusters in Vertically** Faris Alqadah, Raj Bhatnagar (*University of Cincinnati, USA*)

**7D - Industry Day (1)** (Session Chair: Marc Najork)

*Location: Cabernet & Zinfandel*

**The Mountains or the Street Lamp: Search, Research, and Research Again** Christopher J. C. Burges (*Microsoft Research, USA*)

**Crowdsourcing for Relevance Evaluation** Daniel E. Rose (*A9.com, USA*)

**Hadoop: Industrial-Strength Open Source for Data-Intensive Supercomputing** Doug Cutting (*Yahoo!, USA*)

**8A - IR: Enterprise Search** (Session Chair: David A. Evans)Location: *Syrah & Merlot***Multi-Aspect Expertise Matching for Review Assignment** Maryam Karimzadehgan, Cheng Xiang Zhai, Geneva Belford (*University of Illinois at Urbana-Champaign, USA*)**Dr. Searcher and Mr. Browser: A Unified Hyperlink-Click Graph** Barbara Poblete, Carlos Castillo, Aristides Gionis (*University Pompeu Fabra and Yahoo! Research Barcelona, Spain*)**Modeling Multi-step Relevance Propagation for Expert Finding** Pavel Serdyukov, Henning Rode, Djoerd Hiemstra (*University of Twente, The Netherlands*)**Trada: Tree Based Ranking Function Adaptation** Keke Chen, Rongqing Lu, CK Wong, Gordon Sun, Larry Heck, Belle Tseng (*Yahoo!, USA*)**8B - IR: Structured Documents** (Session Chair: Mário J. Gaspar da Silva)Location: *Stags Leap & Howell Mountain***Structural Relevance: A Common Basis for the Evaluation of Structured Document Retrieval** Sadek Ali, Mariano Consens, Gabriella Kazai, Mounia Lalmas (*University of Toronto, Canada, Microsoft Research Cambridge and Queen Mary University of London, UK*)**A Generative Retrieval Model for Structured Documents** Le Zhao, Jamie Callan (*Carnegie Mellon University, USA*)**A Densitometric Approach to Web Page Segmentation** Christian Kohlschütter, Wolfgang Nejdl (*University of Hannover, Germany*)**Using Structured Text for Large-Scale Attribute Extraction** Sujith Ravi, Marius Pasca (*University of Southern California and Google, Inc., USA*)**8C - KM: Text Mining** (Session Chair: Alessandro Moschitti)Location: *Mt. Veeder & Carneros***Identification of Class Specific Discourse Patterns** Anup Kumar Chalamalla, Sumit Negi, L.Venkata Subramaniam, Ganesh Ramakrishnan (*IBM India Research Lab and Indian Institute of Technology, India*)**Scalable Community Discovery on Textual Data with Relations** Huajing Li, Zaiqing Nie, Wang-Chien Lee, C. Lee Giles, Ji-Rong Wen (*Pennsylvania State University, USA and Microsoft Research Asia, China*)**Information Shared by Many Objects** Chong Long, Xiaoyan Zhu, Ming Li, Bin Ma (*Tsinghua University, China*)**Extremely Fast Text Feature Extraction for Classification and Indexing** George Forman, Evan Kirshenbaum (*HP Labs, USA*)**8D - Industry Day (2)** (Session Chair: Bruce Croft)Location: *Cabernet & Zinfandel***Statistical Learning as the Ultimate Agile Development Tool** Peter Norvig (*Google, USA*)**The Evolving Computational Advertising Landscape** Andrei Broder (*Yahoo! Research, USA*)

**9A - DB: Mobile and Distributed Data Management** (Session Chair: Talel Abdessalem)Location: *Mt. Veeder & Carneros***Valid Scope Computation for Location-Dependent Spatial Query in Mobile Broadcast Environments** Ken C. K. Lee, Josh Schiffman, Baihua Zheng, Wang-Chien Lee (*Pennsylvania State University, USA* and *Singapore Management University, Singapore*)**Adaptive Distributed Indexing for Structured Peer-to-Peer Networks** Linh Nguyen, Wai Gen Yee, Ophir Frieder (*Illinois Institute of Technology, USA*)**PROQID: Partial Restarts of Queries in Distributed Databases** Jon Olav Hauglid, Kjetil Norvag (*Norwegian University of Science and Technology, Norway*)**9B - IR: Question Answering** (Session Chair: Lucian Vlad Lita)Location: *Stags Leap & Howell Mountain***Answering Questions with Authority** Andrew Hickl (*Language Computer Corporation, USA*)**Cache-aware Load Balancing for Question Answering** David Dominguez-Sal, Mihai Surdeanu, Josep Aguilar-Saborit, Josep-LL. Larriba-Pey (*DAMA-UPC, Spain*)**A System for Finding Biological Entities that Satisfy Certain Conditions from Texts** Wei Zhou, Clement Yu, Weiyi Meng (*University of Illinois at Chicago* and *SUNY Binghamton, USA*)**9C - KM: Information Extraction** (Session Chair: Eugene Agichtein)Location: *Syrah & Merlot***Intra-Document Structural Frequency Features for Semi-Supervised Domain Adaptation** Andrew Arnold, William Cohen (*Carnegie Mellon University, USA*)**Academic Conference Homepage Understanding Using Constrained Hierarchical Conditional Random Fields** Xin Xin, Juanzi Li, Jie Tang, Qiong Luo (*Tsinghua University* and *HKUST, China*)**Identifying Table Boundaries in Digital Documents via Sparse Line Detection** Ying Liu, Prasenjit Mitra, C. Lee Giles (*Pennsylvania State University, USA*)**9D - Industry Day (3)** (Session Chair: Andrew Tomkins)Location: *Cabernet & Zinfandel***Toward Next Generation Search: Business, Product, Science, Infrastructure, and Talent** William Chang (*Baidu.com, China*)**Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO** Ronny Kohavi (*Microsoft, USA*)**The Secret History of Silicon Valley** Steven Gary Blank (*Stanford University, USA*)

# *Poster Rules & Best Poster*

## **When to be there**

This year we will have three poster sessions, one for each day of the main conference. Each day will have about 35 posters organized by area track. On each day the poster session will be held from 9:00 am. to 5:30 pm. The poster days will be: Monday (October 27, 2008); Tuesday (October 28, 2008); Wednesday (October 29, 2008).

Each day there will be a poster boaster session that will form part of the morning plenary session; here the area chairs will present a synopsis of each poster for that day.

## **Best Poster Award**

A best poster award will be given to the authors of the best poster. The best poster will be determined as follows: Each conference participant will receive pocket money, \$100 in fake money, as part of their registration package. Each poster will be assigned a “tip-jar” (an envelope), where conference participants can deposit tips, i.e., part of their pocket money. At the end of the poster day, each poster author turns in his/her tip-jar to the registration desk. The author must insure that the money is counted and the total entered on the envelope; envelopes with no totals will be assign a total of zero. The five posters with the most “tips” will be considered for the best poster award. One will be selected by the best poster committee as the best poster. The best poster award will be presented during the closing ceremony of the main conference.

## **Set up**

On the day of your poster please setup your poster by 9:00 am; the poster area will be open from 8:00 am onwards.

## **Tear down**

Please take down your posters after 5:30 pm.

## **Who to contact with questions**

A student assistant will be available in the poster room. They should be easily identifiable by the CIKM 2008 t-shirt that they all wear. For more complex issues, email [info@cikm2008.org](mailto:info@cikm2008.org) and [cikm2008pc-chairs@soe.ucsc.edu](mailto:cikm2008pc-chairs@soe.ucsc.edu) with the following subject “CIKM 2008: Posters Questions.”

**DATABASE**

- 1.01 Privacy-Preserving Data Publishing for Horizontally Partitioned Databases**  
Pawel Jurczyk, Li Xiong (*Emory University, USA*)
- 1.02 CE2--Towards a Large Scale Hybrid Search Engine with Integrated Ranking Support**  
Haofen Wang, Thanh Tran (*Shanghai Jiao Tong University, China and University Karlsruhe, Germany*)
- 1.03 Scaling up Duplicate Detection in Graph Data**  
Melanie Herschel, Felix Naumann (*Hasso-Plattner-Institut, Germany*)
- 1.04 ROAD: An Efficient Framework for Location Dependent Spatial Queries on Road Networks**  
Ken C. K. Lee, Wang-Chien Lee, Bailhua Zheng (*Pennsylvania State University, USA and Singapore Management University, Singapore*)
- 1.05 View and Index Selection for Query-Performance Improvement: Quality-Centered Algorithms and Heuristics**  
Maxim Kormilitsin, Rada Chirkova, Yahya Fathi, Matthias Stallmann (*North Carolina State University, USA*)
- 1.06 SQL Extension for Exploring Multiple Tables**  
Sung Jin Kim, Junghoo (John) Cho (*University of California, Los Angeles, USA*)
- 1.07 PBFilter: Indexing Flash-Resident Data through Partitioned Summaries**  
Shaoyi Yin, Philippe Pucheral, Xiaofeng Meng (*INRIA, France and PRISM & Renmin University of China, China*)

**INDUSTRY**

- 1.08 Transaction Reordering with Application to Synchronized Scans**  
Gang Luo, Jeffrey Naughton, Curt Ellmann, Michael Watzke (*IBM T.J. Watson Research Center, USA*)
- 1.09 Yizkor Books: A Voice for the Silent Past**  
Jason Soo, Rebecca Cathey, Ophir Frieder, Michlean Amir, Gideon Frieder (*Illinois Institute of Technology and BAE Systems, USA*)

**INFORMATION RETRIEVAL**

- 1.10 An Approximate String Matching Approach for Handling Incorrectly Typed URLs**  
Mihai Stroe, Radu Berinde, Cosmin Negruseri, Dan Popovici (*Google Switzerland GmbH, Switzerland*)
- 1.11 Speed up Semantic Search in P2P Networks**  
Wang Qiang, Rui Li, Lei Chen, Jie Lian, Tamer Ozsü (*University of Waterloo, Canada and HKUST, China*)
- 1.12 A Note Search Based Forecasting of Ad Volume in Contextual Advertising**  
Xuerui Wang, Andrei Broder, Marcus Fontoura, Vanja Josifovski (*University of Massachusetts, Amherst, Yahoo! Research, USA and PUC-Rio, Brazil*)
- 1.13 An Extension of PLSA for Document Clustering**  
Young Min Kim, Jean-François Pessiot, Massih Amini, Patrick Gallinari (*Pierre and Marie Curie University, France*)
- 1.14 Online Spam-Blog Detection Through Blog Search**  
Linhong Zhu, Aixin Sun, Byron Choi (*Nanyang Technological University, Singapore*)
- 1.15 Nested Region Algebra Extended with Variables for Tag-Annotated Text Search**  
Katsuya Masuda, Junichi Tsujii (*University of Tokyo, Japan and University of Manchester, UK*)
- 1.16 Searching the Wikipedia with Contextual Information**  
Antti Ukkonen, Carlos Castillo, Debora Donato, Aristides Gionis (*Yahoo! Research Barcelona, Spain*)
- 1.17 Winoing-Based Text Clustering**  
Javier Parapar, Álvaro Barreiro (*University of A Coruña, Spain*)
- 1.18 Using Sequence Classification for Filtering Web Pages**  
Binyamin Rosenfeld, Ronen Feldman, Lyle Ungar (*Hebrew University, Israel and University of Pennsylvania, USA*)
- 1.19 Passage Relevance Models for Genomics Search**  
Jay Urbain, Ophir Frieder, Nazli Goharian (*Illinois Institute of Technology, USA*)
- 1.20 Cross-Document Cross-Lingual Coreference Retrieval**  
Elif Aktolga, Marc-Allen Cartright, James Allan (*University of Massachusetts, Amherst, USA*)
- 1.21 Siphon++: A Hidden-WebCrawler for Keyword-Based Interfaces**  
Karane Vieira, Luciano Barbosa, Juliana Freire, Altigran Silva (*University of Utah, USA and Federal University of Amazonas, Brazil*)
- 1.22 Investigating External Corpus and Clickthrough Statistics for Query Expansion in the Legal Domain**  
Tonya Custis, Khalid Al-Kofahi (*Thomson Reuters, USA*)
- 1.23 Corpus Microsurgery: Criteria Optimization for Medical Cross-Language IR**  
Monica Rogati, Yiming Yang, Jaime Carbonell (*LinkedIn and Carnegie Mellon University, USA*)
- 1.24 Metadata Extraction and Indexing for Map Search in Web Documents**  
Tan Qingzhao, Prasenjit Mitra, C. Lee Giles (*Pennsylvania State University, USA*)

**KNOWLEDGE MANAGEMENT**

- 1.25 An Integration Strategy for Mining Product Features and Opinions**  
Qingliang Miao, Qiudan Li, Ruwei Dai (*CAS, China*)
- 1.26 Overlapping Community Structure Detection in Networks**  
Nan Du, Bin Wu, Bai Wang (*Beijing University of Posts and Telecom, China*)
- 1.27 Coreference Resolution using Expressive Logic Models**

Ki Chan, Wai Lam, Xiaofeng Yu (*The Chinese University of Hong Kong, China*)

**1.28 A Method to Predict Social Annotations**

Ming-Hung Hsu, Hsin-Hsi Chen (*National Taiwan University, Taiwan*)

**1.29 Large Maximal Cliques Enumeration in Sparse Graphs**

Natwar Modani, Kuntal Dey (*IBM India Research Lab, India*)

**1.30 Summarization of Social Activity over Time: People, Actions and Concepts in Dynamic Networks**

Yu-Ru Lin, Hari Sundaram, Aisling Kelliher (*Arizona State University, USA*)

**1.31 Using Tag Semantic Network for Keyphrase Extraction in Blogs**

Lizhen Qu, Iryna Gurevych, Christof Müller (*Technische Universität Darmstadt, Germany*)

**1.32 Handling Implicit Geographic Evidence for Geographic IR**

Nuno Cardoso, Mário J. Gaspar da Silva, Diana Santos (*University of Lisbon, Portugal*)

**1.33 Estimating Real-valued Characteristics of Criminals from their Recorded Crimes**

Richard Bache, Fabio Crestani (*University of Strathclyde, UK and University of Lugano, Switzerland*)

**1.34 Representative Entry Selection for Profiling Blogs**

Jinfeng Zhuang, Steven C.H. Hoi, Aixin Sun, Rong Jin, Maxim Kormilitsin (*Nanyang Technological University, Singapore and Michigan State University, USA*)

**1.35 Efficient Web Matrix Processing based on Dual Reordering**

Hsu Chih-Ming, Ming-Syan Chen (*National Taiwan University, Taiwan*)

**1.36 CoreEx: Heuristic Content Extraction from Online News Articles**

Jyotika Prasad, Andreas Paepcke (*Stanford University, USA*)

**1.37 A Novel Email Abstraction Scheme for Spam Detection**

Chi-Yao Tseng, Ming-Syan Chen, Pin-Chieh Sung (*National Taiwan University, Taiwan*)

**1.38 Tag-Based Filtering for Personalized Bookmark Recommendations**

Pavan Kumar Vatturi, Werner Geyer, Casey Dugan, Michael Muller, Beth Brownholtz (*Oregon State University, USA*)

**1.39 Closing the Loop in Webpage Understanding**

Chunyu Yang, Yong Cao, Zaiqing Nie, Jie Zhou, Ji-Rong Wen (*Microsoft Research Asia, China*)



**DATABASE**

- 2.01 Efficient Processing of Probabilistic Spatio-Temporal Range Queries over Moving Objects**  
Bruce Chung, Wang-Chien Lee, Arbee Chen (*Pennsylvania State University, USA*)
- 2.02 Data Degradation: Making Private Data Less Sensitive Over Time**  
Nicolas Ancaix, Luc Bouganim, Harold J. W. van Heerde, Philippe Pucheral, Peter M. G. Apers (*INRIA, France and University of Twente, The Netherlands*)
- 2.03 A Light Weighted Damage Tracking Quarantine and Recovery Scheme for Mission-Critical Database Systems**  
Kun Bai (*Pennsylvania State University, USA*)
- 2.04 Query Optimization in XML-Based Information Integration**  
Dongfeng Chen, Rada Chirkova, Maxim Kormilitsin, Fereidoon Sadri, Timo Salo (*North Carolina State University and University of North Carolina Greensboro, USA*)
- 2.05 Estimating the Number of Answers with Guarantees for Structured Queries in P2P Databases**  
Marcel Karnstedt, Kai-Uwe Sattler, Michael Haß, Manfred Hauswirth, Brahmananda Sapkota, Roman Schmidt (*TU Ilmenau, Germany*)
- 2.06 Evaluating Partial Tree-Pattern Queries on XML Streams**  
Xiaoying Wu, Dimitri Theodoratos (*New Jersey Institute of Technology, USA*)
- 2.07 Characterization of TPC-H Queries for a Column-Oriented Database on a Dual-Core AMD Athlon Processor**  
Pranav Vaidya, Jaehwan John Lee (*Indiana University, Purdue University Indianapolis, USA*)

**INFORMATION RETRIEVAL**

- 2.08 Natural Language Retrieval of Grocery Products**  
Petteri Nurmi, Emil Lagerspetz, Wray Buntine, Patrik Floreen, Joonas Kukkonen, Peter Peltonen (*Helsinki Institute for Information Technology, Finland*)
- 2.09 Improve the Effectiveness of the Opinion Retrieval and Opinion Polarity Classification**  
Wei Zhang, Lifeng Jia, Clement Yu, Weiyi Meng (*University of Illinois at Chicago and SUNY Binghamton, USA*)
- 2.10 A Latent Variable Model for Query Expansion Using the Hidden Markov Model**  
Qiang Huang, Dawei Song (*The Open University, UK*)
- 2.11 A Survey of Pre-Retrieval Query Performance Predictors**  
Claudia Hauff, Djoerd Hiemstra, Franciska de Jong (*University of Twente, The Netherlands*)
- 2.12 Modeling Document Features for Expert Finding**  
Jianhan Zhu, Dawei Song, Stefan Rueger, Jimmy Huang (*University College London, The Open University, Imperial College London, UK and York University, Canada*)
- 2.13 Mining Named Entity Transliteration Equivalents from Comparable Corpora**  
Raghavendra Udapa, Saravanan K. Kumaran, A. Jagadeesh Jagarlamudi (*Microsoft Research, India*)
- 2.14 Estimating Retrieval Effectiveness using Rank Distributions**  
Vishwa Vinay, Ingemar Cox, Natasa Milic-Frayling (*Microsoft Research Labs, Cambridge, UK*)
- 2.15 Semi-supervised Ranking Aggregation**  
Shouchun Chen, Fei Wang, Yaangqiu Song, Changshui Zhang (*Tsinghua University, China*)
- 2.16 Ranking in Folksonomy Systems: Can Context Help?**  
Fabian Abel, Nicola Henze, Daniel Krause (*Leibniz University, Germany*)
- 2.17 Evaluating Topic Models for Information Retrieval**  
Xing Yi, James Allan (*University of Massachusetts Amherst, USA*)
- 2.18 A Novel Statistical Chinese Language Model and Its Application in Pinyin-to-Character Conversion**  
Bo Lin, Jun Zhang (*NTU, Singapore*)
- 2.19 Integrating Clustering and Multi-Document Summarization to Improve Document Understanding**  
Dingding Wang, Shenghuo Zhu, Yun Chi, Tao Li (*Florida International University and NEC Laboratories America, USA*)
- 2.20 Answering General Time Sensitive Queries**  
Wisam Dakka, Luis Gravano, Panagiotis Ipeirotis (*Columbia University and New York University, USA*)
- 2.21 Search-based Query Suggestion**  
Jiang-Ming Yang, Rui Cai, Feng Jing, Shuo Wang, Lei Zhang, Wei-Ying Ma (*Microsoft Research Asia, China*)
- 2.22 Entity-Based Query Reformulation Using Wikipedia**  
Yang Xu, Fan Ding, Bin Wang (*Institute of Computing Technology, China*)

**KNOWLEDGE MANAGEMENT**

- 2.23 Group-based Learning -- A Boosting Approach**  
Weijian Ni, Jun Xu, Hang Li, Yalou Huang (*Nankai University and Microsoft Research Asia, China*)
- 2.24 Collaborative Partitioning with Maximum User Satisfaction**  
Fred Annexstein, Svetlana Strunjas (*University of Cincinnati, USA*)
- 2.25 Efficient Frequent Pattern Mining over Data Streams**  
Syed Tanbeer, Chowdhury Ahmed, Byeong-Soo Jeong, Young-Koo Lee (*Kyung Hee University, Korea*)
- 2.26 GHOST: An Effective Graph-based Framework for Name Distinction**  
Xiaoming Fan, Jianyong Wang, Bing Lv, Lizhu Zhou, Wei Hu (*Tsinghua University, China*)

**2.27 Deriving Non-Redundant Approximate Association Rules from Hierarchical Datasets**

Gavin Shaw, Yue Xu, Shlomo Geva (*Queensland University of Technology, Australia*)

**2.28 Pattern-based Semantic Class Discovery with Multi-Membership Support**

Shuming Shi, Xiaokang Liu, Ji-Rong Wen (*Microsoft Research Asia, China*)

**2.29 Detecting Significant Distinguishing Sets Among Bi-clusters**

Faris Alqadah, Raj Bhatnagar (*University of Cincinnati, USA*)

**2.30 Semi-Supervised Metric Learning by Maximizing Constraint Margin**

Fei Wang, Shouchun Chen, Tao Li, Changshui Zhang (*Tsinghua University, China and Florida International University, USA*)

**2.31 On Quantifying Changes in Temporally Evolving Dataset**

Rohan Choudhary, Sameep Mehta, Amitabha Bagchi (*Indian Institute of Technology, India*)

**2.32 Fast Spatial Co-location Mining Without Cliques Checking**

Zhongshan Lin, Seungjin Lim (*Utah State University, USA*)

**2.33 Decomposition of Terminology Graphs for Domain Knowledge Acquisition**

Fidelia Ibekwe-SanJuan, Eric SanJuan, Michael Vogeley (*University of Lyon, France and Drexel University, USA*)

**2.34 In the Development of a Spanish MetaMap**

Francisco Carrero, José Carlos Cortizo, José María Gómez Hidalgo, Manuel de Buenaga (*Universidad Europea de Madrid, Spain*)

**2.35 Scalable Complex Pattern Search in Sequential Data**

Kaghazian Leila, Reza Sadri, Dennis McLeod (*University of Southern California, USA*)

**2.36 Combining Concept Hierarchies and Statistical Topic Models**

Chaitanya Chemudugunta, Padhraic Smyth, Mark Steyvers (*University of California, Irvine, USA*)

**DATABASE**

- 3.01 Energy-Efficient Skyline Query Processing and Maintenance in Sensor Networks**  
Weifa Liang, Baichen Chen, Jeffrey Xu (*Australian National University, Australia and Chinese University of Hong Kong, China*)
- 3.02 Table Summarization with the Help of Domain Lattices**  
K. Selcuk Candan, Huiping Cao, Yan Qi, Maria Luisa Sapino (*Arizona State University, USA*)
- 3.03 Protecting Location Privacy against Location-Dependent Attack in Mobile Services**  
Xiao Pan, Jianliang Xu, Xiaofeng Meng (*RUC, PR. and Hong Kong Baptist University, China*)
- 3.04 Polyhedral Transformation for Indexed Rank Order Correlation Queries**  
Philon Nguyen, Nematollaah Shiri (*Concordia University, USA*)
- 3.05 Workload-Based Optimization of Integration Processes**  
Matthias Boehm, Dirk Habich, Wolfgang Lehner, Uwe Wloka (*Dresden University of Applied Sciences, Germany*)

**INFORMATION RETRIEVAL**

- 3.06 Re-Considering Collaborative Filtering Parameters in the Context of New Data**  
Adele Howe, Ryan Forbes (*Colorado State University, USA*)
- 3.07 Efficient Estimation of the Size of Text Deep Web Data Source**  
Jianguo Lu (*University of Windsor, Canada*)
- 3.08 A GeoReferencing Multistage Method for Locating Geographic Context in Web Search**  
Álvaro Zubizarreta, Pablo de la Fuente, Jose M. Cantera, Mario Arias, Jorge Cabrero, Guido García, César Llamas, Jesús Vegas (*Universidad de Valladolid, Spain*)
- 3.09 Suppressing Outliers in Pairwise Preference Ranking**  
Vitor Carvalho, Jonathan Elsas, William Cohen, Jaime Carbonell (*Carnegie Mellon University, USA*)
- 3.10 Incorporating Place Name Extents into Geo-IR Ranking**  
Hiroyuki Toda, Norihito Yasuda, Yumiko Matsuura, Ryoji Kataoka (*NTT Corporation, Japan*)
- 3.11 The Effect of Contextualization at Different Granularity Levels in Content-oriented XML Retrieval**  
Paavo Arvola, Jaana Kekalainen, Marko Junkkari (*University of Tampere, Finland*)
- 3.12 Using Current Browsing Context to Improve Search Relevance**  
Mandar Rahurkar, Silviu Cucerzan (*University of Illinois at Urbana Champaign and Microsoft Research, USA*)
- 3.13 Using a Graph-based Ontological User Profile for Personalizing Search**  
Mariam Daoud, Lynda Tamine-Lechani, Mohand Boughanem (*IRIT, France*)
- 3.14 Measuring User Preference Changes in Digital Libraries**  
Yang Sun, Huajing Li, Isaac G. Councill, Wang-Chien Lee, C. Lee Giles (*Pennsylvania State University, USA*)
- 3.15 Utilization of Navigational Queries for Result Presentation and Caching in Search Engines**  
Rifat Ozcan, Ismail Altinoglu, Özgür Ulusoy (*Bilkent University, Turkey*)
- 3.16 ShopSmart: Making Recommendations based on Technical Specifications and User Reviews**  
Alexander Yates, James Joseph, Alexander Cohn, Nick Sillick, Ana-Maria Popescu (*Temple University, USA*)
- 3.17 Trust, Authority and Popularity in Social Information Retrieval**  
Gabriella Kazai, Natasa Milic-Frayling (*Microsoft Research, UK*)
- 3.18 A Spam Resistant Family of Concavo-Convex Ranks for Link Analysis**  
Sreangsu Acharyya, Joydeep Ghosh (*University of Texas Austin, USA*)

**KNOWLEDGE MANAGEMENT**

- 3.19 Boosting Social Annotations Using Propagation**  
Shenghua Bao, Bohai Yang, Ben Fei, Shengliang Xu, Zhong Su, Yong Yu (*Shanghai Jiao Tong University, China*)
- 3.20 Effective Pattern Taxonomy Mining in Text Documents**  
Yuefeng Li, Sheng-Tang Wu, Xiaohui Tao (*Queensland University of Technology, Australia*)
- 3.21 Incorporating Topical Support Documents into a Small Training Set in Text Categorization**  
Kyung Soon Lee (*Chonbuk National University, Korea*)
- 3.22 Exploiting Context to Detect Sensitive Information in Call Center Conversations**  
Tanveer Faruque, Sumit Negi, Anup Kumar Chalamalla, L. Venkata Subramaniam (*IBM India Research Laboratory, India*)
- 3.23 Multi-scale Characterization of Social Network Dynamics in the Blogosphere**  
Munmun De Choudhury, Hari Sundaram, Ajita John, Doree Duncan Seligmann (*Arizona State University, USA*)
- 3.24 Semi-supervised Text Categorization by Active Search**  
Zenglin Xu, Rong Jin, Kaizhu Huang, Michael R. Lyu, Irwin King (*Chinese University Hong Kong, China and Michigan State University, USA*)
- 3.25 Clustering Multi-way data via Adaptive Subspace Iteration**  
Wei Peng, Tao Li, Bo Shao (*Florida International University, USA*)
- 3.26 A Coarse-grain Grid-based Subspace Clustering Method for Online Multi-dimensional Data Streams**  
Jae Woo Lee, Won Suk Lee (*Yonsei University, Korea*)
- 3.27 A Matrix-based Approach for Semi-supervised Document Co-clustering**  
Yanhua Chen, Lijun Wang, Ming Dong (*Wayne State University, USA*)
- 3.28 Categorizing Bloggers' Interests Based on Short Snippets of Blog Posts**  
Jiahui Liu, Larry Birnbaum, Bryan Pardo (*Northwestern University, USA*)

# Industry Day

October 26-30, 2008 Napa Valley, California  
Wednesday, October 29, 2008

**Organizers:** Marius Pasca (*Google, USA*) and James G. Shanahan (*Church and Duncan Group, Inc., USA*)

## Session 7D

Session Chair: Marc Najork  
10:15-11:45 – Cabernet & Zinfandel

### **The Mountains or The Street Lamp: Search, Research, and Research Again**

Christopher J. C. Burges (*Microsoft Research, USA*)

### **Crowdsourcing for Relevance Evaluation**

Daniel E. Rose (*A9.com, USA*)

### **Hadoop: Industrial-Strength Open Source for Data-Intensive Supercomputing**

Doug Cutting (*Yahoo!, USA*)

## Session 8D

Session Chair: Bruce Croft  
13:30-15:00 – Cabernet & Zinfandel

### **Statistical Learning as the Ultimate Agile Development Tool**

Peter Norvig (*Google, USA*)

### **The Evolving Computational Advertising Landscape**

Andrei Z. Broder (*Yahoo! Research, USA*)

## Session 9D

Session Chair: Andrew Tomkins  
15:30-17:00 – Cabernet & Zinfandel

### **Toward Next Generation Search: Business, Product, Science, Infrastructure, and Talent**

William I. Chang (*Baidu.com, China*)

### **Practical Guide to Controlled Experiments on the Web: Listen to Your Customers not to the HiPPO**

Ronny Kohavi (*Microsoft, USA*)

### **The Secret History of Silicon Valley**

Steven Gary Blank (*Stanford University and University of California Berkeley*)

**Panel #1: E-Discovery**

Tuesday, Oct 28, 10:15-11:45

Location: Cabernet &amp; Zinfandel

Chair: David A. Evans

It is common practice in the U.S. for courts to require that the parties to a legal case make available to one another all material relevant to the case, including electronically held data and documents. For large corporations, such relevant information may encompass terabytes of e-mail and other files spanning many years. The challenge of E-Discovery in response to a court order is (in a relatively short amount of time) to identify, assemble, individuate, access, categorize, and analyze an organization's electronically held material, segregate all "privileged" material (which can be legally withheld), and present to the court all (and only) the required documents. The techniques needed to accomplish such a task necessarily include search, clustering, classification, filtering, social network analysis, extraction, and more - and no one of these is sufficient. The panel will describe this problem in detail, providing additional background and context on E-Discovery generally, and will explore specific techniques and cases that amply demonstrate why E-Discovery is quintessentially a "CIKM" problem - multi-disciplinary, multi-technology, and "multi-difficult."

**Why E-Discovery is a CIKM-Hard Problem**David A. Evans (*JustSystems Evans Research, USA*)**Panning for Gold in E-Discovery: What Every Information Scientist Should Know About the Way Lawyers Search for Electronic Evidence**Jason R. Baron (*National Archives and Records Administration, USA*)**IR Perspectives on the E-Discovery Problem**Chris Buckley (*Sabir Research, USA*)**Technical Cases Studies from the E-Discovery Front Lines**Robert S. Bauer (*H5, USA*)**Panel #2: The Social (Open) Workspace**

15:30-17:00

Location: Cabernet &amp; Zinfandel

Chair: David A. Evans

Social networking promises individuals new dimensions of freedom to interact, associate, and give expression to their talents. Recently, systems such as Mechanical Turk have started to facilitate self-organizing collaboration on work-related tasks. Such developments raise interesting questions. Is it possible to create (and sustain) businesses that do not have traditional, formal structure - without traditional "employees?" Can we find and organize (and optimize) talent on the web for task-oriented work - spontaneously and efficiently? How do people relate to one another in possibly evanescent workgroups? One aspect of the challenge in the Social Workspace is understanding and modeling the user behavior and the economic basis for creating, preserving, and exchanging value in the marketplace when workgroup identity, orientations to property, recruiting and managing appropriate talent are not organized under traditional company structures. Another aspect is the technology needed to support virtual organizations and work. The panel will discuss trends in social work and the evolving (scientific) basis of our understanding of new models of workers and organizations.

**Automating Knowledge**Susan Feldman (*IDC, USA*)**How to Augment Social Cognition**Ed H. Chi (*Palo Alto Research Center (PARC), USA*)**Trust and Design in Social Systems**Natasa Milic-Frayling (*Microsoft Research (Cambridge), USA*)**Using Social Networks for Social Work**Igor Perisic (*LinkedIn, USA*)

**Full Day Workshops**

10:00 – 11:00: Coffee Break

Lunch Varies by Workshop

15:30 – 16:00: Coffee Break

**BooksOnline 2008 Setting the Roadmap for Research Advances in Large Digital Book Repositories**

Location: Carneros

Time: 9:00 – 17:00

This workshop aims to bring together researchers and industry practitioners, in Information Retrieval, Digital Libraries, eBooks, Human Computer Interaction, Publishing industry and on-line book services, to map out the agenda for digital book research, producing a draft research agenda for the next 5 years.

Paul Kantor (*Rutgers LAIR Laboratory, USA*), Gabriella Kazai (*Microsoft Research, UK*), Natasa Milic-Frayling (*Microsoft Research, UK*), Ross Wilkinson (*CSIRO, Australia*)

Webpage: <http://research.microsoft.com/workshops/BooksOnline08>

**DTMBIO 2008 ACM Second International Workshop on Data and Text Mining in Bioinformatics**

Location: Castle Peak

Time: 8:30 – 17:30

The focus of this workshop is to bring together researchers from data and text mining and computational biology that want to integrate heterogeneous structured and unstructured data, and to better embed literature information into bioinformatics solutions.

Min Song (*New Jersey Institute of Technology, USA*), Luke Huan (*University of Kansas, USA*), Doheon Lee (*KAIST, Korea*)

Webpage: <http://web.njit.edu/~song/dtmbio2008/home.html>

**iNEWS 2008 Improving Non-English Web Searching**

Location: Zinfandel

Time: 8:00 – 17:00

The main aims of the second edition of this workshop are to identify the problems in web searching when working with non-English languages and to propose solutions and tools so as to improve the existing situation.

Fotis Lazarinis (*Technological Educational Institute of Mesolonghi, Greece*), Efthimis Efthimiadis (*University of Washington, USA*), Jesus Vilares (*University of A Coruna, Spain*), John Tait (*Information Retrieval Facility, Austria*)

Webpage: <http://rea.teimes.gr/lazarinf/iNEWS08/>

**SSM 2008 Search in Social Media**

Location: Merlot

Time: 9:00 – 17:30

Social applications are the fastest growing segment of the web. Social media establish new forums for content creation, and permit novel applications at the intersection of people and information. The purpose of this workshop is to focus the attention of the research community on search in particular kinds of social media, such as blogs, and forums such as Facebook, Myspace, Flickr, and the like.

Ian Soboroff (*National Institute of Standards and Technology, USA*), Eugene Agichtein (*Emory University, USA*), Ravi Kumar (*Yahoo! Research, USA*)

Webpage: <http://ir.mathcs.emory.edu/SSM2008/>

**WICOW 2008 Second Workshop on Information Credibility on the Web**

Location: Mt. Veeder

Time: 8:30 – 17:30

The workshop will focus on how to evaluate the credibility of information on the Web. How can we extract credible information related to a given topic, organize this information, detect its provenance, clarify the background, and identify facts, and various related opinions.

Katsumi Tanaka (*Kyoto University, Japan*), Takashi Matsuyama (*Kyoto University, Japan*), Ee-Peng Lim (*Nanyang University, Singapore*) Adam Jatowt (*Kyoto University, Japan*)

Webpage: <http://www.dl.kuis.kyoto-u.ac.jp/wicow2/>

**WIDM 2008 Tenth International Workshop on Web Information and Data Management**

Location: Syrah

Time: 8:45 – 17:45

The objective of the workshop is to bring together researchers, industrial practitioners, and developers to study how Web information can be extracted, stored, analyzed, and processed to provide useful knowledge to the end users for various advanced database and Web applications.

Chee-Yong Chan (*National University of Singapore, Singapore*), Neoklis Polyzotis (*University of California-Santa Cruz, USA*)

Webpage: <http://widm2008.comp.nus.edu.sg/>

**GIR 2008 Fifth Workshop on Geographic Information Retrieval**

Location: Diamond Mountain

Time: 7:00 – 12:45

There is a need for specialised geographical information techniques that exhibit geographical intelligence in recognising the presence of place names, and other geographic references, in documents and user queries and provide retrieval facilities that can make imprecise matches between geographical query terms and the contents of documents. This workshop is intended to cover a wide range of topics that are of particular concern to geographical information retrieval.

Ross Purves (*University of Zurich, Switzerland*), Chris Jones (*Cardiff University, UK*)Webpage: <http://www.geo.unizh.ch/~rsp/gir08/>**LSDS-IR 2008 Workshop on Large-Scale Distributed Systems for Information Retrieval**

Location: Cabernet

Time: 8:00 – 12:25

Scalability problems in information retrieval have to be addressed in the near future, and new distributed applications are likely to drive the way in which people use the Web. Distributed IR is the point in which these two directions converge. This workshop will provide space for researchers to discuss these problems and to define new directions for the work on distributed information retrieval.

Sebastian Michel (*Ecole Polytechnique Federale de Lausanne, Switzerland*), Gleb Skobeltsyn (*Ecole Polytechnique Federale de Lausanne, Switzerland*), Wai Gen Yee (*Illinois Institute of Technology, USA*)Webpage: <http://lsirwww.epfl.ch/LSDS-IR08/>**PIKM 2008 PhD workshop**

Location: Atlas Peak

Time: 7:00 – 12:45

The goal of this workshop is to encourage Ph.D. students to present their dissertation research at a relatively early stage. The targeted students are those with a focus in one or more of the CIKM research tracks, i.e., databases, information retrieval and knowledge management. This workshop will enable them to present their Ph.D. dissertation proposals and/or ongoing research on their dissertation sub-problems.

Aparna Varde (*Virginia State University, USA*), Prasan Roy (*Aster Data Systems, USA*)Webpage: <http://web.cs.wpi.edu/~aparna/PIKM2008CFP.htm>**Afternoon Workshops****DOLAP 2008 ACM Eleventh International Workshop on Data Warehousing and OLAP**

Location: Diamond Mountain

Time: 13:00 – 19:00

DOLAP will focus on new research directions and emerging application domains in the areas of Data Warehousing and Online Analytical Processing (OLAP) technologies.

Alberto Abello (*Polytechnical University of Catalonia, Spain*), Il-Yeon Song (*Drexel University, USA*)Webpage: <http://www.cis.drexel.edu/faculty/song/dolap/DOLAP08/DOLAP08-CfP-03.htm>**ONISW 2008 Second International workshop on Ontologies and Information Systems for the Semantic Web**

Location: Atlas Peak

Time: 13:00 – 19:00

The main objective of the workshop is to bring together researchers in Information Management interested in the relation between ontology and information models, to present results and to discuss theoretical aspects and good practice.

Ramez Elmasri (*University of Texas at Arlington, USA*), Mathias Brochhausen (*Institute of Formal Ontology, Germany*), Martin Doerr (*Foundation for Research and Technology, Greece*), Hyoil Han (*Drexel University, USA*)Webpage: <http://www.ischool.drexel.edu/faculty/hhan/onisw2008/>**PaIR 2008 Patent Information Retrieval**

Location: Cabernet

Time: 13:00 – 18:45

The objective of the workshop is to provide a forum for Information Retrieval and Knowledge Management scientists as well as Patent Retrieval experts from industry to study the next generation of patent search tools.

Helmut Berger (*Matrixware Information Services GmbH, Austria*), Michael Dittenbach (*Matrixware Information Services GmbH, Austria*), Georg Sommer (*Information Retrieval Facility, Austria*), John Tait (*Information Retrieval Facility, Austria*)Webpage: <http://www.ir-facility.org/events/pair08>

# Conference Venue

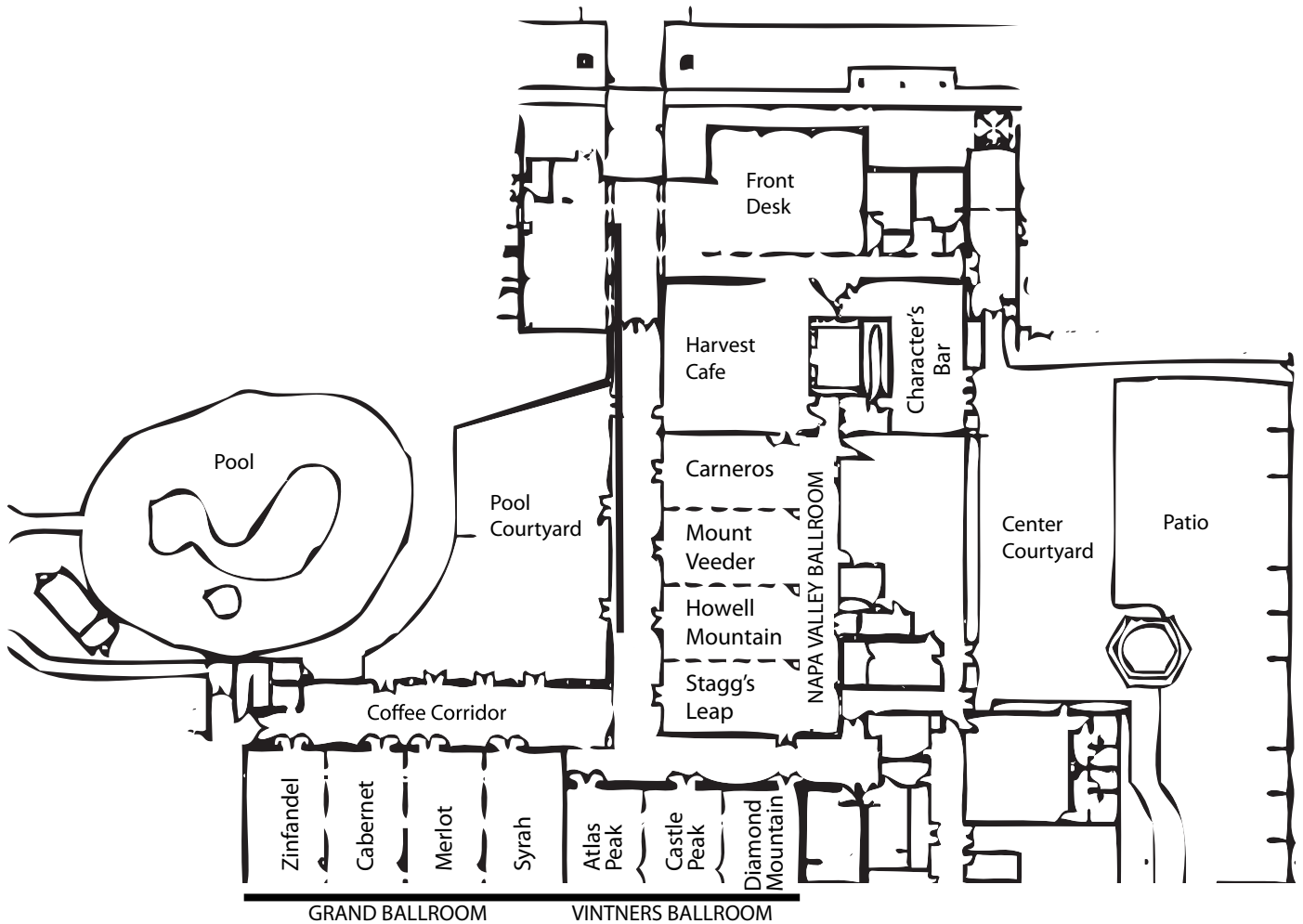
## Internet Access

Sponsored by Platform A, CIKM attendees can access the internet via several conveniently placed wall jacks or WiFi. The connection details are as follows:

ssid: cikh2008  
 user: cikh2008  
 password: platform\_a

If you require assistance with the connection, please call:  
 (800) 487-2090.

Room/Section	Ballroom
Syrah	Grand Ballroom
Merlot	Grand Ballroom
Cabernet	Grand Ballroom
Zinfandel	Grand Ballroom
Stags Leap	Napa Valley Ballroom
Howell Mountain	Napa Valley Ballroom
Mt. Veeder	Napa Valley Ballroom
Carneros	Napa Valley Ballroom
Diamond Mtn	Vintners Ballroom
Castle Peak	Vintners Ballroom
Atlas Peak	Vintners Ballroom





# Banquet

## Overview

On October 28, 2008, join us for the CIKM 2008 Banquet at COPIA. COPIA: The American Center for Wine, Food and the Arts will furnish CIKM attendees and guests with a Halloween-themed banquet in the heart of Napa Valley. The evening will begin with a champagne toast and formal reception, followed by a seated three-course meal at the world-renowned Julia's Kitchen Restaurant. Since it is Halloween, **costumes and masks are encouraged**, as attendees unwind, relax, and enjoy the company of each other with good food and fine wine.

**Important Note: No admittance granted without ticket and conference badge. Guests must be accompanied by a conference attendee. Absolutely no exceptions. Thanks for your cooperation!**

## Schedule

17:30-18:45	Shuttle buses continually (every 15 minutes) depart Marriott hotel parking lot for COPIA. Arriving attendees will be directed to the North Garden
17:45-19:00	Welcome Reception - beverages and refreshments.
19:00-22:45	Dinner
19:15	Welcome Speech
20:15	Announcements & Awards
19:15-22:30	Buses to run every 15 minutes back to Marriott hotel

**COPIA -The American Center for Wine, Food and the Arts**  
 500 1st St.  
 Napa, CA 94559  
[www.copia.org](http://www.copia.org)

### About Copia

Founded by Robert Mondavi, Julia Child and other luminaries of American wine and food, COPIA is the leading national organization dedicated to consumer wine and food appreciation. Located in the heart of Napa Valley, COPIA is the premier wine country destination—offering visitors exceptional wine and food-tasting programs, exhibitions, organic edible gardens, films, concerts, fine and casual dining and shopping.

### About Julia's Kitchen Restaurant

Named for culinary legend Julia Child, Julia's Kitchen has been touted by the New York Times and placed among the top three restaurants in the Napa Valley.

