CIKM'10 Workshops and Tutorials at a Glance

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

Tuesday, October 26, 2010

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

Saturday, October 30, 2010

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

Workshop Programs

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

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

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Program Schedule:

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

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

Discussion *– 10 min*

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

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

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

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

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

Discussion *– 10 min*

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

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

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

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

Discussion *– 10 min*

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

Discussion *– 10 min*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Third International Workshop on Patent Information Retrieval (PaIR 2010)

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

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Second International Workshop on Cloud Data Management (CloudDB 2010)

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

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

Ordonez, Carlos

Dynamic Data Replication through Virtualization
Sergey Savinov, Khuzaima Daudjee

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

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

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

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

Saturday, October 30, 2010 (Room: Algonquin)

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Program Schedule:

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

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

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

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

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

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

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

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

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

Semantic Annotations in Clinical Documentation
Sumithra Velupillai

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

Is Formalizing Events Necessary for Full Exploitation
Antonio Badia

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

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

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

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

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

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

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

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

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

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

Modeling Betweenness for Question Answering
Brandeis Marshall

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

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

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

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

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

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

Saturday, October 30, 2010 (Room: Alberta)

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

Entity Classification by Bag of Wikipedia Articles
Tomas Kliegr

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

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

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

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

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

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

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

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

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

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

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

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

Modeling a Trust Cloud Context
Sean Thorpe

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

Saturday, October 30, 2010 (Room: Manitoba)

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

Program Schedule:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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