

Tuesday, 20th July

Tuesday, 9:00-10:15 Keynote address

- **Refactoring the Search Problem**
Gary W. Flake (*Microsoft Research Redmond*)

Tuesday, 10:15-10:45 Break

Tuesday, 10:45-12:00

Session 1A: Clustering I

Chair: Gabriella Pasi (*University of Milano-Bicocca*)

- **Prototype Hierarchy Based Clustering for Web Collection Categorization and Navigation**
Zhao-Yan Ming, Kai Wang, Tat-Seng Chua (*National University of Singapore*)
- **Person Name Disambiguation by Bootstrapping**
Yoshida Minoru, Ikeda Masaki, Ono Shingo, Sato Issei, Nakagawa Hiroshi (*University of Tokyo*)
- **Self-Taught Hashing for Fast Similarity Search**
Dell Zhang (*Birkbeck, University of London*), Jun Wang (*University College London*),
Deng Cai (*Zhejiang University*), Jinsong Lu (*Birkbeck, University of London*)

Session 1B: User Models

Chair: Ian Ruthven (*University of Strathclyde*)

- **Personalizing Information Retrieval for Multi-Session Tasks: The Roles of Task Stage and Task Type**
Jingjing Liu, Nicholas J. Belkin (*Rutgers, The State University of New Jersey*)
- **Predicting Searcher Frustration**
Henry Feild, James Allan (*University of Massachusetts Amherst*), Rosie Jones (*Yahoo! Labs*)
- **The Good, the Bad, and the Random: An Eye-Tracking Study of Ad Quality in Web Search**
Georg Buscher (*University of Kaiserslautern*), Susan Dumais (*Microsoft Research Redmond*),
Edward Cutrell (*Microsoft Research India*)

Session 1C: Applications I

Chair: Luo Si (*Purdue University*)

- 103 **Ranking using Multiple Document Types in Desktop Search**
Jinyoung Kim, W. Bruce Croft (*University of Massachusetts Amherst*)
- **Acquisition of Instance Attributes via Labeled and Related Instances**
Enrique Alfonseca, Marius Pasca, Enrique Robledo-Arnuncio (*Google*)
 - **Relevance and Ranking in Online Dating Systems**
Fernando Diaz, Donald Metzler, Sihem Amer-Yahia (*Yahoo! Labs*)

Tuesday, 12:00-14:00 Lunch on your own

Tuesday, 14:00-15:40

Session 2A: Search Engine Architectures and Scalability

Chair: Alistair Moffat (*University of Melbourne*)

- **Scalability of Findability: Effective and Efficient IR Operations in Large Information Networks**
Weimao Ke, Javed Mostafa (*University of North Carolina at Chapel Hill*)

- **Caching Search Engine Results over Incremental Indices**
Roi Blanco, Edward Bortnikov, Flavio Junqueira, Ronny Lempel (*Yahoo! Research*), Luca Telloi (*Barcelona Supercomputing Center*), Hugo Zaragoza (*Yahoo! Research*)
- **Query Forwarding in Geographically Distributed Search Engines**
B. Barla Cambazoglu (*Yahoo! Research*), Emre Varol, Enver Kayaaslan, Cevdet Aykanat (*Bilkent University*), Ricardo Baeza-Yates (*Yahoo! Research*)
- **A Joint Probabilistic Classification Model for Resource Selection**
Dzung Hong, Luo Si, Paul Bracke, Michael Witt, Tim Juchcinski (*Purdue University*)

Session 2B: Link Analysis & Advertising

Chair: Tie-Yan Liu (*Microsoft Research Asia*)

- **Temporal Click Model for Sponsored Search**
Wanhong Xu (*Carnegie Mellon University*), Eren Manavoglu, Erick Cantu-Paz (*Yahoo! Labs*)
- **Freshness Matters: In Flowers, Food, and Web Authority**
Na Dai, Brian D. Davison (*Lehigh University*)
- **The Importance of Anchor-Text for Ad Hoc Search Revisited**
Marijn Koolen, Jaap Kamps (*University of Amsterdam*)
- **Ready to Buy or Just Browsing? Detecting Web Searcher Goals From Interaction Data**
Qi Guo, Eugene Agichtein (*Emory University*)

Session 2C: Learning to Rank

Chair: Hang Li (*Microsoft Research Asia*)

- **Learning to Efficiently Rank**
Lidan Wang (*University of Maryland, College Park*), Jimmy Lin (*University of Maryland, College Park*), Donald Metzler (*Yahoo!*)
- **Ranking for the Conversion Funnel in Contextual Advertising**
Abraham Bagherjeiran, Andrew Hatch, Adwait Ratnaparkhi (*Yahoo!*)
- **How Good is a Span of Terms? Exploiting Proximity to Improve Web Retrieval**
Krysta Svore (*Microsoft Research*), Pallika Kanani (*University of Massachusetts Amherst*), Nazan Khan (*Microsoft Research*)
- **Learning to Rank Only Using Training Data From Related Domain**
Wei Gao (*The Chinese University of Hong Kong*), Peng Cai (*East China Normal University*), Kam-Fai Wong (*The Chinese University of Hong Kong*), Aoying Zhou (*East China Normal University*)

Tuesday, 15:40-16:15 Break

Tuesday, 16:15-17:30

Session 3A: Clustering II

Chair: Omar Alonso (*Microsoft/Bing.com*)

- **Optimal Meta Search Results Clustering**
Claudio Carpineto, Giovanni Romano (*Fondazione Ugo Bordoni, Rome*)
- **Analysis of Structural Relationships for Hierarchical Cluster Labeling**
Markus Muhr, Roman Kern (*Know-Center GmbH*), Michael Granitzer (*University of Technology Graz*)
- **On the Existence of Obstinate Results in Vector Space Models**
Miloš Radovanović (*University of Novi Sad*), Alexandros Nanopoulos (*University of Hildesheim*), Mirjana K. Ivanović (*University of Novi Sad*)

Session 3B: Filtering and Recommendation

Chair: Douglas W. Oard (*University of Maryland, College Park*)

- **Social Media Recommendation Based on People and Tags**
Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, Erel Uziel (*IBM Research Haifa*)
- **A Network-Based Model for High-Dimensional Information Filtering**
Nikolaos Nanas, Manolis Vavalis (*Centre for Research and Technology Thessaly*),
Anne De Roeck (*The Open University*)
- **Temporal Diversity in Recommender Systems**
Neal Lathia, Stephen Hailes, Licia Capra (*University College London*),
Xavier Amatriain (*Telefonica Research*)
- **Serendipitous Recommendation via Innovators**
N. N.

Session 3C: Information Retrieval Theory

Chair: Hugo Zaragoza (*Yahoo! Research*)

- **On Statistical Analysis and Optimization of Information Retrieval Effectiveness Metrics**
Jun Wang, Jianhan Zhu (*University College London*)
- **Information-Based Models for Ad Hoc Information Retrieval**
Stéphane Clinchant (*Xerox Research Centre Europe, Université Grenoble 1*),
Eric Gaussier (*Université Joseph Fourier*)
- **Score Distribution Models: Assumptions, Intuition, and Robustness to Score Manipulation**
Evangelos Kanoulas (*University of Sheffield*),
Keshi Dai, Virgil Pavlu, Javed A. Aslam (*Northeastern University*)

Tuesday, 17:30-19:00 Poster & Demonstrations reception

Wednesday, 21st July

Wednesday, 9:00-10:15 Keynote address

- **Is the Cranfield Paradigm Outdated?**
Donna Harman (*NIST*)

Wednesday, 10:15-10:45 Break

Wednesday, 10:45-12:00

Session 4A: Language Models & IR Theory

Chair: James Allan (*University of Massachusetts Amherst*)

- **Geometric Representations for Multiple Documents**
Jangwon Seo, W. Bruce Croft (*University of Massachusetts Amherst*)
- **Using Statistical Decision Theory and Relevance Models for Query-Performance Prediction**
Anna Shtok, Oren Kurland (*Technion - Israel Institute of Technology*),
David Carmel (*IBM Research*)
- **Active Learning for Ranking Through Expected Loss Optimization**
Bo Long, Olivier Chapelle, Ya Zhang, Yi Chang, Zhaohui Zheng, Belle Tseng (*Yahoo! Labs*)

Session 4B: Query Representations & Reformulations

Chair: Maarten de Rijke (*University of Amsterdam*)

- **Image Search by Concept Map**
Hao Xu (*University of Science and Technology of China*),
Jingdong Wang, Xian-Sheng Hua, Shipeng Li (*Microsoft Research Asia*)

- **Generalized Syntactic and Semantic Models of Query Reformulation**
Amac Herdagdelen (*University of Trento*), Massimiliano Ciaramita, Daniel Mahler (*Google*), Maria Holmqvist (*Linköping University*), Keith Hall, Stefan Riezler, Enrique Alfonseca (*Google*)
- **Evaluating Verbose Query Processing Techniques**
Samuel Huston, W. Bruce Croft (*University of Massachusetts Amherst*)

Session 4C: Automatic Classification

Chair: Eric Gaussier (*Université Joseph Fourier*)

- **SED: Supervised Experimental Design and Its Application to Text Classification**
Yi Zhen, Dit-Yan Yeung (*Hong Kong University of Science and Technology*)
- **Temporally-Aware Algorithms for Document Classification**
Thiago Salles (*Universidade Federal de Minas Gerais*), Leonardo Chaves Dutra Rocha (*Universidade Federal de São João Del-Rei*), Gisele L. Pappa, Fernando Mourão, Wagner Meira Jr., Marcos André Gonçalves (*Universidade Federal de Minas Gerais*)
- **Multi-Label Classification with Meta-Level Features**
Siddharth Gopal, Yiming Yang (*Carnegie Mellon University*)

Wednesday, 12:00-14:00 Lunch on your own

Wednesday, 14:00-15:40

Session 5A: Retrieval Models and Ranking

Chair: Djoerd Hiemstra (*University of Twente*)

- **Estimation of Statistical Translation Models Based on Mutual Information for Ad Hoc Information Retrieval**
Maryam Karimzadehgan, ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)
- **DivQ: Diversification for Keyword Search over Structured Databases**
Elena Demidova (*L3S Research Center, Leibniz Universität Hannover*), Peter Fankhauser (*L3S Research Center*), Xuan Zhou (*CSIRO ICT Centre*), Wolfgang Nejdl (*L3S Research Center*)
- **Finding Support Sentences for Entities**
Roi Blanco, Hugo Zaragoza (*Yahoo! Research*)
- **Estimating Probabilities for Effective Data Fusion**
David Lillis, Lusheng Zhang, Fergus Toolan, Rem Collier, John Dunnion (*University College Dublin*)

Session 5B: User Feedback & User Models

Chair: Nicholas J. Belkin (*Rutgers, The State University of New Jersey*)

- **Incorporating Post-Click Behaviors Into a Click Model**
Feimin Zhong, Dong Wang (*Tsinghua University*), Gang Wang, Weizhu Chen, Yuchen Zhang, Zheng Chen, Haixun Wang (*Microsoft Research Asia*)
- **Interactive Retrieval Based on Faceted Feedback**
Lanbo Zhang, Yi Zhang (*University of California at Santa Cruz*)
- **A Comparison of General vs. Personalized Affective Models for the Prediction of Topical Relevance**
Ioannis Arapakis, Konstantinos Athanasakos, Joemon M. Jose (*University of Glasgow*)
- **Understanding Web Browsing Behaviors through Weibull Analysis of Dwell Time**
Chao Liu, Ryan White, Susan Dumais (*Microsoft Research Redmond*)

Session 5C: Web IR and Social Media Search

Chair: Iadh Ounis (*University of Glasgow*)

- **Segmentation of Multi-Sentence Questions: Towards Effective Question Retrieval in cQA Services**
Kai Wang, Zhao-Yan Ming, Xia Hu, Tat-Seng Chua (*National University of Singapore*)
- **Mining the Blogosphere for Top News Stories Identification**
Yeha Lee, Hun-Young Jung, Woosang Song, Jong-Hyeok Lee (*POSTECH, Pohang*)
- **Proximity-Based Opinion Retrieval**
Shima Gerani, Mark J. Carman, Fabio Crestani (*University of Lugano*)
- **Evaluating and Predicting Answer Quality in Community QA**
Chirag Shah, Jeffrey Pomerantz (*University of North Carolina at Chapel Hill*)

Wednesday, 15:40-16:15 Break

Wednesday, 16:15-17:30

Session 6A: Document Structure & Adversarial Information Retrieval

Chair: Mounia Lalmas (*University of Glasgow*)

- **Adaptive Near-Duplicate Detection via Similarity Learning**
Hannaneh Hajishirzi (*University of Illinois at Urbana-Champaign*),
Wen-Tau Yih, Aleksander Kolcz (*Microsoft Research*)
- **A Content Based Approach for Discovering Missing Anchor Text for Web Search**
Xing Yi, James Allan (*University of Massachusetts Amherst*)
- **Uncovering Social Spammers: Social Honeypots + Machine Learning**
Kyumin Lee, James Caverlee (*Texas A&M University*),
Steve Webb (*Georgia Institute of Technology*)

Session 6B: Users and Interactive IR

Chair: David Carmel (*IBM Research Haifa*)

- **Studying Trailfinding Algorithms for Enhanced Web Search**
Adish Singla (*Microsoft Bing*), Ryen White (*Microsoft Research Redmond*),
Jeff Huang (*University of Washington*)
- **Context-Aware Ranking in Web Search**
Biao Xiang (*University of Science and Technology of China*), Daxin Jiang (*Microsoft Research Asia*),
Jian Pei (*Simon Fraser University*), Xiaohui Sun (*Microsoft Research Asia*),
Enhong Chen (*University of Science and Technology of China*),
Hang Li (*Microsoft Research Asia*)
- **Collecting High Quality Overlapping Labels at Low Cost**
Hui Yang (*Carnegie Mellon University*), Anton Mityagin (*Qualcomm*),
Krysta Svore (*Microsoft Research*)

Session 6C: Document Representation and Content Analysis

Chair: Marie-Francine Moens (*Katholieke Universiteit Leuven*)

- **Multi-Style Language Model for Web Scale Information Retrieval**
Kuansan Wang, Jianfeng Gao, Xiaolong Li (*Microsoft Research*)
- **Combining Coregularization and Consensus-Based Self-Training for Multilingual Text Categorization**
Massih-Reza Amini (*National Research Council of Canada*),
Cyril Goutte (*NRC Institute for Information Technology*),
Nicolas Usunier (*University Pierre et Marie Curie*)
- **Towards Subjectifying Text Clustering**
Sajib Dasgupta, Vincent Ng (*University of Texas at Dallas*)

Wednesday, 19:00-22:00 Conference Banquet

Thursday, 22nd July

Thursday, 9:00-10:15

Session 7A: Test-Collections

Chair: John Tait (*The Information Retrieval Facility*)

- **The Effect of Assessor Error on IR System Evaluation**
Ben Carterette (*University of Delaware*), Ian Soboroff (*NIST*)
- **Building Reusable Test Collections Through Experimental Design**
Ben Carterette (*University of Delaware*), Evangelos Kanoulas (*University of Sheffield*),
Virgil Pavlu (*Northeastern University*), Hui Fang (*University of Delaware*)
- **Do User Preferences and Evaluation measures Line Up?**
Mark Sanderson, Monica Lestari Paramita, Paul Clough,
Evangelos Kanoulas (*University of Sheffield*)

Session 7B: Query Log Analysis

Chair: Yoelle Maarek (*Yahoo! Research*)

- **Query Similarity by Projecting the QueryFlow Graph**
Ilaria Bordino (*Sapienza University of Rome & University of Pompeu Fabra*)
Carlos Castillo, Debora Donato, Aristides Gionis (*Yahoo! Research Barcelona*)
- **The Demographics of Web Search**
Ingmar Weber, Carlos Castillo (*Yahoo! Research Barcelona*)
- **A User Behavior Model for Average Precision and its Generalization to Graded Judgments**
Georges Dupret (*Yahoo! Labs*), Benjamin Piwowarski (*University of Glasgow*)

Thursday, 10:15-10:45 Break

Thursday, 10:45-12:00

Session 8A: Summarization & User Feedback

Chair: Elizabeth D. Liddy (*Syracuse University*)

- **Learning More Powerful Statistical Tests for Click-Based Retrieval Evaluation**
Yisong Yue, Yue Gao (*Cornell University*), Olivier Chapelle, Ya Zhang (*Yahoo! Labs*),
Thorsten Joachims (*Cornell University*)
- **EUSUM: Extracting Easy-to-Understand English Summaries for Non-Native Readers**
Xiaojun Wan (*Peking University*)
- **Visual Summarization of Web Pages**
Binxing Jiao (*University of Science and Technology of China*),
Linjun Yang, Xu Jizheng, Feng Wu (*Microsoft Research Asia*)

Session 8B: Query Analysis

Chair: Ricardo Baeza-Yates (*Yahoo! Research*)

- **Estimating Advertisability of Tail Queries for Sponsored Search**
Sandeep Pandey, Kunal Punera, Marcus Fontoura, Vanja Josifovski (*Yahoo! Research*)
- **Exploring Reductions for Long Web Queries**
Niranjan Balasubramanian (*University of Massachusetts Amherst*),
Giridhar Kumaran, Vitor Carvalho (*Microsoft Corporation*)
- **Positional Relevance Model for Pseudo-Relevance Feedback**
Yuanhua Lv, ChengXiang Zhai (*University of Illinois at Urbana-Champaign*)

Thursday, 12:00-14:00 Business Lunch

Thursday, 14:00-15:40

Session 9A: Effectiveness Measures

Chair: Ian Soboroff (*NIST*)

- **Assessing the Scenic Route: Measuring the Value of Search Trails in Web Logs**
Ryen W. White (*Microsoft Research Redmond*), Jeff Huang (*University of Washington*)
- **Human Performance and Retrieval Precision Revisited**
Mark D. Smucker, Chandra Prakash Jethani (*University of Waterloo*)
- **Extending Average Precision to Graded Relevance Judgments**
Stephen Robertson (*Microsoft Research*), Evangelos Kanoulas (*University of Sheffield*),
Emine Yilmaz (*Microsoft Research Cambridge*)
- **PRES: A Score Metric for Evaluating Recall-Oriented Information Retrieval Applications**
Walid Magdy, Gareth J.F. Jones (*Dublin City University*)

Session 9B: Multimedia Information Retrieval

Chair: Tat Seng Chua (*National University of Singapore*)

- **Content-Enriched Classifier for Web Video Classification**
Ce Zhang, Bin Cui (*Peking University*), Gao Cong (*Nanyang Technological University*)
- **Robust Audio Identification for MP3 Popular Music**
Li Wei, Liu Yaduo, Xue Xiangyang (*Fudan University, Shanghai*)
- **Effective Music Tagging Through Advanced Statistical Modeling**
Jialie Shen (*Singapore Management University*), Meng Wang (*Microsoft Research Asia*),
Shuicheng Yan, HweeHwa Pang (*Singapore Management University*),
Xian-Sheng Hua (*Microsoft Research Asia*)
- **Properties of Optimally Weighted Data Fusion in CBMIR**
Peter Wilkins, Alan Smeaton, Paul Ferguson (*Dublin City University*)

Thursday, 15:40-16:15 Break

Thursday, 16:15-17:30

Session 10A: Non-English IR & Evaluation

Chair: Jaana Kekäläinen (*University of Tampere*)

- **To Translate or Not to Translate?**
Chia-Jung Lee, Chin-Hui Chen, Shao-Hang Kao, Pu-Jen Cheng (*National Taiwan University*)
- **Multilingual PRF: English Lends a Helping Hand**
Manoj Kumar Chinnakotla, Karthik Raman, Pushpak Bhattacharyya (*IIT Bombay*)
- **Comparing the Sensitivity of Information Retrieval Metrics**
Filip Radlinski (*Microsoft, Cambridge*), Nick Craswell (*Microsoft, Redmond*)

Session 10B: Applications II

Chair: David Lewis (*David D. Lewis Consulting*)

- **Efficient Partial-Duplicate Detection Based on Sequence Matching**
Qi Zhang, Yue Zhang, Haomin Yu, Xuanjing Huang (*Fudan University, Shanghai*)
- **Discriminative Models of Integrating Document Evidence and Document-Candidate Associations for Expert Search**
Yi Fang, Luo Si, Aditya Mathur (*Purdue University*)
- **Vertical Selection in the Presence of Unlabeled Verticals**
Jaime Arguello (*Carnegie Mellon University*),
Fernando Diaz, Jean-Francois Paiement (*Yahoo! Labs*)

Thursday, 17:30-17:45

Closing Ceremony