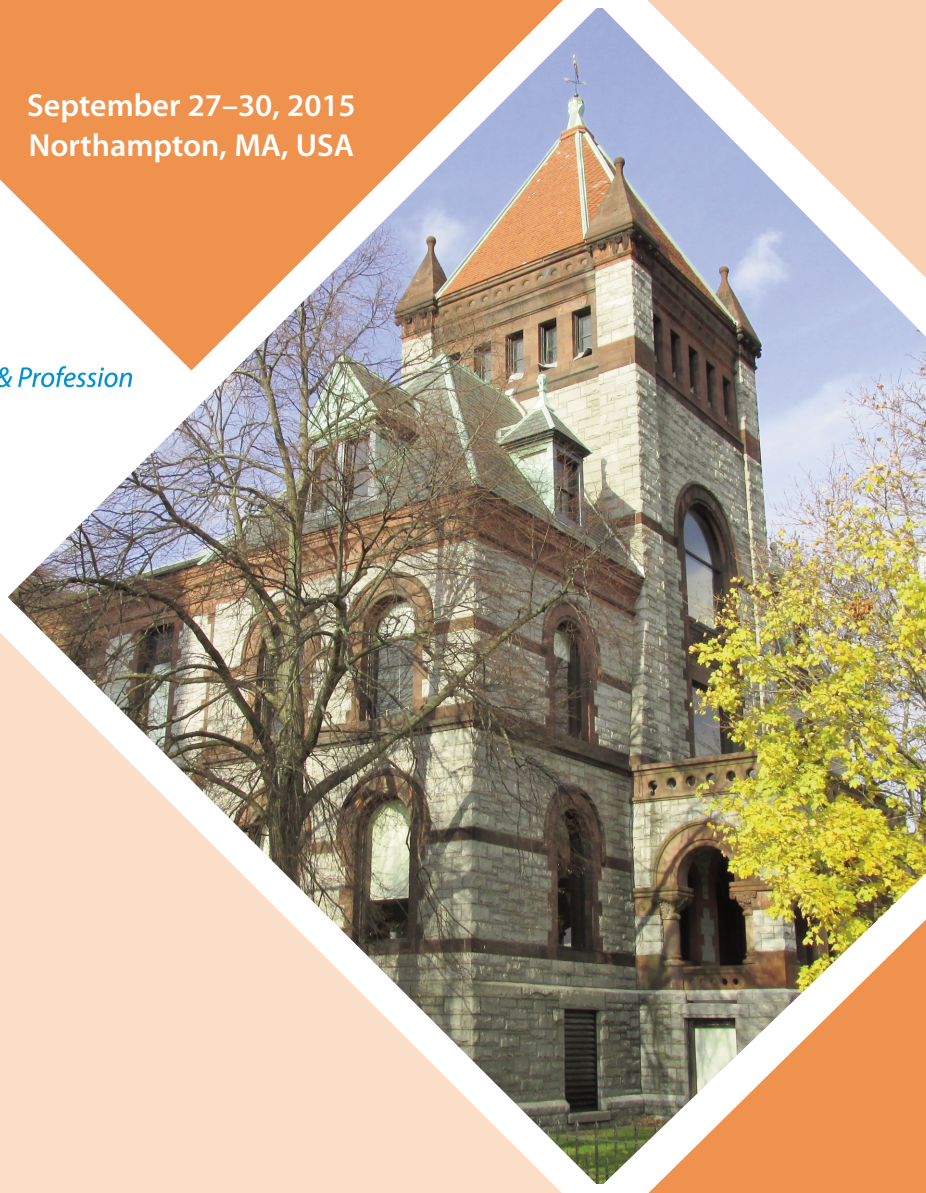
September 27–30, 2015
Northampton, MA, USA



Association for Computing Machinery

Advancing Computing as a Science & Profession

ICTIR'15

Proceedings of the 2015 ACM SIGIR International Conference on the Theory of Information Retrieval

Sponsored by:

ACM SIGIR

Supported by:

Google, Microsoft Research, Oracle Labs, and Lexalytics

Association for
Computing Machinery

Advancing Computing as a Science & Profession

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York, New York 10121-0701

Copyright © 2015 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: permissions@acm.org or Fax +1 (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through www.copyright.com.

Notice to Past Authors of ACM-Published Articles
ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that has been previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ISBN: 978-1-4503-3833-2

Additional copies may be ordered prepaid from:

ACM Order Department
PO Box 30777
New York, NY 10087-0777, USA

Phone: 1-800-342-6626 (USA and Canada)
+1-212-626-0500 (Global)
Fax: +1-212-944-1318
E-mail: acmhelp@acm.org
Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA

ICTIR 2015 General Chairs' Welcome

Welcome to ICTIR 2015, the 5th International Conference on the Theory of Information Retrieval and the first conference with that name to be fully sponsored by the ACM Special Interest Group on Information Retrieval (SIGIR). Becoming part of the SIGIR family, along with the well-known SIGIR conference and the new CHIIR conference, will improve both the coordination between these venues and the availability of resources such as student scholarships. We believe that ICTIR is a natural complement to the SIGIR conference, with a stronger emphasis on theory and formal approaches, and less emphasis on large-scale experimentation and practice.

Given that ICTIR has generally been a smaller, more discussion-oriented conference than SIGIR, we felt that the Pioneer Valley, and Northampton in particular, would be an ideal venue for a September meeting. We hope to provide a relaxing small-town setting, beautiful Fall colors, and enough restaurants and bars to satisfy attendees used to living in large cities. The Center for Intelligent Information Retrieval and the University of Massachusetts are just on the other side of the Connecticut River in Amherst, and Noah Webster and Melvil Dewey did some of their best work at Amherst College, so there is a strong association with information retrieval. Given the quality of research that will be presented at ICTIR 2015, this should be another milestone in this local IR history. Enjoy the conference!

James Allan
College of Information and Computer Sciences
University of Massachusetts
Amherst

Bruce Croft
College of Information and Computer Sciences
University of Massachusetts
Amherst

Program Chairs' Welcome

The program chairs are excited to welcome you to the ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR); a new episode in the history of ICTIR where the conference has been embraced by ACM SIGIR and now runs for the first time under their flag.

The conference aims to provide a forum for the presentation and discussion of research related to the *foundational* aspects of Information Retrieval (IR), including, for example, new or improved models of relevance, ranking, representation, information needs, and evaluation. Norbert Fuhr helped define foundation precisely as *a scientific result that others can build upon and use for their own research*. The class of scientific works solicited in the call for papers thus covers both purely theoretical contributions as well as reproducible experimental results; where reproducibility is emphasized (e.g., through sharing the data sets and/or software implementations).

The conference has been very successful in attracting high quality contributions, helping us compose a strong and varied program. In response to the call for contributions, the conference received a healthy number of submissions: 4 tutorials out of which we accepted 2 (or 50%), 43 short paper submissions where we accepted 22 (51%), and, 57 long paper submissions out of which 29 have been included in the final program (51%). (As an aside, due to the high quality of the submission pool, we would like to emphasize that acceptance rate is obviously not an indicator of quality!)

We thank all the program committee members for their effort and excellent reviews, the Tutorial Chair, Peter Bruza, and his committee for the selection of tutorials, and the General Chairs, James Allan and Bruce Croft, for their guidance. We invite you to enjoy the scientific program with exciting novel results in our field!

Arjen de Vries and ChengXiang Zhai
(Long paper PC Chairs)

Norbert Fuhr and Yi Zhang
(Short paper PC Chairs)

Table of Contents

ICTIR 2015 Organization ix

ICTIR 2015 Sponsors & Supporters xii

Keynote & Tutorials

- **Embedded Representations of Lexical and Knowledge-Base Semantics** 1
  Andrew McCallum *(University of Massachusetts, Amherst)*
- **Theory of Retrieval: The Retrievability of Information** 3
  Leif Azzopardi *(University of Glasgow)*
- **Statistical Significance Testing in Information Retrieval: Theory and Practice** 7
  Ben Carterette *(University of Delaware)*

Session M1: IR Evaluation – Measures

- **An Axiomatically Derived Measure for the Evaluation of Classification Algorithms** 11
  Fabrizio Sebastiani *(Qatar Computing Research Institute)*
- **Towards a Formal Framework for Utility-oriented Measurements of Retrieval Effectiveness** 21
  Marco Ferrante, Nicola Ferro, Maria Maistro *(University of Padua)*
- **Bayesian Inference for Information Retrieval Evaluation** 31
  Ben Carterette *(University of Delaware)*

Session M2: Theoretical Framework and Transfer Learning

- **Transferring Learning to Rank Models for Web Search** 41
  Craig Macdonald *(University of Glasgow)*, B. Taner Dinçer *(Mugla University)*, Iadh Ounis *(University of Glasgow)*
- **The Probability Ranking Principle is Not Optimal in Adversarial Retrieval Settings** 51
  Ran Ben Basat, Moshe Tennenholtz, Oren Kurland *(Technion)*
- **Dynamic Information Retrieval: Theoretical Framework and Application** 61
  Marc Sloan, Jun Wang *(University College London)*
- **A Theoretical Analysis of Two-Stage Recommendation for Cold-Start Collaborative Filtering** 71
  Xiaoxue Zhao, Jun Wang *(University College London)*

Session T1: IR Evaluation – Interactive and Task Evaluation

- **An Analysis of Theories of Search and Search Behavior** 81
  Leif Azzopardi *(University of Glasgow)*, Guido Zuccon *(Queensland University of Technology (QUT))*
- **Dynamic Test Collections for Retrieval Evaluation** 91
  Ben Carterette, Ashraf Bah, Mustafa Zengin *(University of Delaware)*
- **Development and Evaluation of Search Tasks for IIR Experiments using a Cognitive Complexity Framework** 101
  Diane Kelly, Jaime Arguello, Ashlee Edwards, Wan-ching Wu *(University of North Carolina at Chapel Hill)*

Session T2: Query Representation

- **Query Expansion with Freebase** 111
  Chenyan Xiong, Jamie Callan *(Carnegie Mellon University)*

- **Language-independent Query Representation for IR Model Parameter Estimation on Unlabeled Collections** 121
  Parantapa Goswami, Massih-Reza Amini, Eric Gaussier *(Université Grenoble Alps)*
- **Terms, Topics & Tasks: Enhanced User Modelling for Better Personalization** 131
  Rishabh Mehrotra, Emine Yilmaz *(University College London)*

Session T3: Language Models for IR

- **Axiomatic Analysis of Smoothing Methods in Language Models for Pseudo-Relevance Feedback** 141
  Hussein Hazimeh, ChengXiang Zhai *(University of Illinois at Urbana-Champaign)*
- **On Divergence Measures and Static Index Pruning** 151
  Ruey-Cheng Chen *(RMIT University)*, Chia-Jung Lee, W. Bruce Croft *(University of Massachusetts, Amherst)*
- **Partially Labeled Supervised Topic Models for Retrieving Similar Questions in CQA Forums** 161
  Debasis Ganguly, Gareth J.F. Jones *(Dublin City University)*

Session T4: Entity Analysis

- **Entity Linking in Queries: Tasks and Evaluation** 171
  Faegheh Hasibi *(Norwegian University of Science and Technology)*, Krisztian Balog *(University of Stavanger)*, Svein Erik Bratsberg *(Norwegian University of Science and Technology)*
- **Random Walks on the Reputation Graph** 181
  Sabir Ribas *(Universidade Federal de Minas Gerais)*, Berthier Ribeiro-Neto *(Universidade Federal de Minas Gerais & Google Inc)*, Rodrygo L. T. Santos *(Universidade Federal de Minas Gerais)*, Edmundo de Souza e Silva *(Universidade Federal do Rio de Janeiro)*, Alberto Ueda *(Universidade Federal de Minas Gerais)*, Nivio Ziviani *(Universidade Federal de Minas Gerais & Zunnit Tech)*
- **Entropy and Graph Based Modelling of Document Coherence Using Discourse Entities: An Application to IR** 191
  Casper Petersen, Christina Lioma, Jakob Grue Simonsen, Birger Larsen *(Aalborg University*, *Copenhagen)*

Session W1: IR Applications

- **Improving Patent Search by Search Result Diversification** 201
  Youngho Kim, W. Bruce Croft *(University of Massachusetts)*
- **On Microblog Dimensionality and Informativeness: Exploiting Microblogs' Structure and Dimensions for Ad-Hoc Retrieval** 211
  Jesus Alberto Rodriguez Perez, Joemon M. Jose *(University of Glasgow)*
- **Online News Tracking for Ad-Hoc Information Needs** 221
  Jeroen B. P. Vuurens *(Delft University of Technology)*, Arjen P. de Vries *(CWI)*, Roi Blanco, Peter Mika *(Yahoo Labs)*

Session W2: NLP and IR

- **IR meets NLP: On the Semantic Similarity between Subject-Verb-Object Phrases** 231
  Dmitrijs Milajevs, Mehrnoosh Sadrzadeh, Thomas Roelleke *(Queen Mary University of London)*
- **A Theoretical Analysis of Cross-lingual Semantic Relatedness in Vector Space Models** 241
  Lei Zhang *(Karlsruhe Institute of Technology (KIT))*, Thanh Tran *(San Jose State University)*, Achim Rettinger *(Karlsruhe Institute of Technology (KIT))*
- **Context Retrieval for Web Tables** 251
  Hong Wang, Anqi Liu, Jing Wang, Brian D. Ziebart, Clement T. Yu *(University of Illinois at Chicago)*, Warren Shen *(Cohesity)*

Session W3: Learning-based Retrieval Models

- **Session Search by Direct Policy Learning** 261
  Jiyun Luo, Xuchu Dong, Hui Yang *(Georgetown University)*
- **Learning to Reinforce Search Effectiveness** 271
  Jiyun Luo, Xuchu Dong, Hui Yang *(Georgetown University)*
- **Learning Asymmetric Co-Relevance** 281
  Fiana Raiber, Oren Kurland *(Technion)*, Filip Radlinski, Milad Shokouhi *(Microsoft)*
- **Implicit Preference Labels for Learning Highly Selective Personalized Rankers** 291
  Paul N. Bennett, Milad Shokouhi, Rich Caruana *(Microsoft)*

Short Papers

- **Anytime Ranking for Impact-Ordered Indexes** 301
  Jimmy Lin *(University of Waterloo)*, Andrew Trotman *(eBay Inc.)*
- **Balancing Aspects in Retrieved Search Results** 305
  David Wemhoener, James Allan *(University of Massachusetts)*
- **Building a Self-Contained Search Engine in the Browser** 309
  Jimmy Lin *(University of Waterloo)*
- **Condensed List Relevance Models** 313
  Fernando Diaz *(Microsoft)*
- **Estimating the Uncertainty of Average F1 Scores** 317
  Dell Zhang *(Birkbeck, University of London)*, Jun Wang, Xiaoxue Zhao *(University College London)*
- **The Feasibility of Brute Force Scans for Real-Time Tweet Search** 321
  Yulu Wang, Jimmy Lin *(University of Waterloo)*
- **Improving Pseudo Relevance Feedback in the Divergence from Randomness Model** 325
  Dipasree Pal, Mandar Mitra *(Indian Statistical Institute)*, Samar Bhattacharya *(Jadavpur University)*
- **An Initial Analytical Exploration of Retrievability** 329
  Aldo Lipani *(National Institute of Informatics)*, Mihai Lupu *(Vienna University of Technology)*,
  Akiko Aizawa *(National Institute of Informatics)*, Allan Hanbury *(Vienna University of Technology)*
- **On the Behavior of PRES Using Incomplete Judgment Sets** 333
  Ellen M. Voorhees *(National Institute of Standards and Technology)*
- **Optimal Packing in Simple-Family Codecs** 337
  Andrew Trotman *(eBay Inc)*, Michael Albert, Blake Burgess *(University of Otago)*
- **Pooling for User-Oriented Evaluation Measures** 341
  Gaurav Baruah, Adam Roegiest, Mark D. Smucker *(University of Waterloo)*
- **Predicting Relevance Feedback Effectiveness with the Help of the Principle of Polyrepresentation in MIR** 345
  David Zellhöfer *(Berlin State Library - Prussian Cultural Heritage)*
- **A Relationship between the Average Precision and the Area Under the ROC Curve** 349
  Wanhua Su *(MacEwan University)*, Yan Yuan *(University of Alberta)*, Mu Zhu *(University of Waterloo)*
- **Revisiting Optimal Rank Aggregation: A Dynamic Programming Approach** 353
  Shayan A. Tabrizi, Javid Dadashkarimi *(University of Tehran)*, Mostafa Dehghani *(University of Amsterdam)*,
  Hassan Nasr Esfahani *(Sharif University of Technology)*, Azadeh Shakery *(University of Tehran)*
- **Searching for Twitter Posts by Location** 357
  Ariana Minot *(Harvard University)*, Andrew Heier, Davis King, Olga Simek,
  Nicholas Stanisha *(MIT Lincoln Laboratory)*
- **A Signaling Game Approach to Databases Querying and Interaction** 361
  Arash Termehchy *(Oregon State University)*, Behrouz Touri *(University of Colorado, Boulder)*
- **Study of Heuristic IR Constraints Under Function Discovery Framework** 365
  Parantapa Goswami, Massih-Reza Amini, Eric Gaussier *(Université Grenoble Alps)*

- **Theoretical Categorization of Query Performance Predictors** 369
  Victor Makarenkov, Bracha Shapira, Lior Rokach *(Ben-Gurion University of the Negev)*
- **Towards Less Biased Web Search** 373
  Xitong Liu, Hui Fang *(University of Delaware)*, Deng Cai *(Zhejiang University)*
- **Two Operators to Define and Manipulate Themes of a Document Collection** 377
  Emanuele Di Buccio, Massimo Melucci *(University of Padua)*
- **Using Part-of-Speech N-grams for Sensitive-Text Classification** 381
  Graham McDonald, Craig Macdonald, Iadh Ounis *(University of Glasgow)*
- **Verboseness Fission for BM25 Document Length Normalization** 385
  Aldo Lipani *(National Institute of Informatics)*, Mihai Lupu, Allan Hanbury *(Vienna University of Technology)*, Akiko Aizawa *(National Institute of Informatics)*

Author Index 389

2015 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR) Organization

General Chairs: James Allan *(University of Massachusetts Amherst, USA)*
Bruce Croft *(University of Massachusetts Amherst, USA)*

Program Chairs: Arjen de Vries *(CWI Amsterdam & Delft University of Technology, The Netherlands)*
ChengXiang Zhai *(University of Illinois at Urbana-Champaign, USA)*

Program Chairs, Short Papers: Norbert Fuhr *(University Duisburg-Essen, Germany)*
Yi Zhang *(University of California, Santa Cruz, USA)*

Tutorial Chair: Peter Bruza *(Queensland University of Technology, Australia)*

Program Committee: Eugene Agichtein *(Yahoo Labs and Emory University, USA)*
Robin Aly *(University of Twente, The Netherlands)*
Giambattista Amati (*Fondazione Ugo Bordoni, Italy*)
Jaime Arguello (*University of North Carolina at Chapel Hill, USA*)
Leif Azzopardi *(University of Glasgow, UK)*
Krisztian Balog *(University of Stavanger, Norway)*
Alejandro Bellogin *(Universidad Autónoma de Madrid, Spain)*
Michael Bendersky *(Google, USA)*
Peter Bruza *(Queensland University of Technology, Australia)*
Katriina Byström *(Oslo and Akershus University College of Applied Sciences, Norway)*
Jamie Callan *(Carnegie Mellon University, USA)*
Mark Carman *(Monash University, Australia)*
Claudio Carpineto *(Fondazione Ugo Bordoni, Italy)*
Ben Carterette *(University of Delaware, USA)*
Pablo Castells *(Universidad Autónoma de Madrid, Spain)*
Yi Chang *(Yahoo! Labs, USA)*
Tat-Seng Chua *(National University of Singapore, Singapore)*
Kevyn Collins-Thompson *(University of Michigan, Ann Arbor, USA)*
Fabio Crestani *(University of Lugano (USI), Switzerland)*
Ronan Cummins *(Cambridge University, UK)*
Jeffrey Dalton *(Google, USA)*
Arjen P. de Vries *(CWI Amsterdam & Delft University of Technology, The Netherlands)*
Gianluca Demartini *(University of Sheffield, UK)*
Thomas Demeester *(Ghent University - iMinds, Belgium)*
Fernando Diaz *(Microsoft, USA)*
Laura Dietz *(Mannheim University, Germany)*

Program Committee (continued):

Miles Efron *(University of Illinois, USA)*
Hui Fang *(University of Delaware, USA)*
Ingo Frommholz *(University of Bedfordshire, UK)*
Norbert Fuhr *(University of Duisburg-Essen, Germany)*
Eric Gaussier *(University J. Fourier/Grenoble 1, France)*
Julio Gonzalo *(ETS de Ingeniería Informática, UNED, Spain)*
Jiafeng Guo *(Chinese Academy of Sciences, China)*
Matthias Hagen *(Bauhaus-Universität Weimar, Germany)*
Allan Hanbury *(TU Wien, Austria)*
Morgan Harvey *(Northumbria University, UK)*
Ben He *(University of Chinese Academy of Sciences, China)*
Jiyin He *(Centrum Wiskunde & Informatica, The Netherlands)*
Djoerd Hiemstra *(University of Twente, The Netherlands)*
Jimmy Huang *(York University, Canada)*
Hideo Joho *(University of Tsukuba, Japan)*
Joemon M Jose *(University of Glasgow, UK)*
Jaap Kamps *(University of Amsterdam, The Netherlands)*
Evangelos Kanoulas *(University of Amsterdam, The Netherlands)*
Diane Kelly *(University of North Carolina at Chapel Hill, USA)*
Alexander Kotov *(Wayne State University, USA)*
Kriste Krstovski *(University of Massachusetts Amherst / Harvard Smithsonian Center for Astrophysics, USA)*
Oren Kurland *(Technion - Israel Institute of Technology, Israel)*
Mounia Lalmas *(Yahoo Labs, UK)*
Wai Lam *(The Chinese University of Hong Kong, Hong Kong)*
Birger Larsen *(Aalborg University, Denmark)*
Victor Lavrenko *(University of Edinburgh, UK)*
Christina Lioma *(University of Copenhagen, Denmark)*
Tie-Yan Liu *(Microsoft Research Asia, China)*
David Losada *(Universidad de Santiago de Compostela, Spain)*
Yuanhua Lv *(Microsoft Research, USA)*
Ilya Markov *(University of Amsterdam, The Netherlands)*
Massimo Melucci *(University of Padova, Italy)*
Donald Metzler *(Google Inc., USA)*
Jian-Yun Nie *(University of Montreal, Canada)*
Gabriella Pasi *(Universita degli studi di Milano Bicocca, Italy)*
Benjamin Piwowarski (*UPMC / CNRS, France)*
Barbara Poblete *(University of Chile, Chile)*
Thomas Roelleke *(Queen Mary University of London, UK)*
Stefan Rueger *(The Open University, UK)*
Tetsuya Sakai (*Waseda University, Japan)*
Mark Sanderson *(RMIT University, Australia)*
Rodrygo Santos *(Universidade Federal de Minas Gerais, Brazil)*

Program Committee (continued):
Mark Smucker *(University of Waterloo, Canada)*
Dawei Song *(The Open University, UK)*
Benno Stein *(University of Weimar, Germany)*
Theodora Tsikrika *(CERTH, Greece)*
Howard Turtle *(Syracuse University, USA)*
Olga Vechtomova *(University of Waterloo, Canada)*
Jun Wang *(University College London, UK)*
Grace Hui Yang *(Georgetown University, USA)*
ChengXiang Zhai *(University of Illinois, USA)*
Yi Zhang *(University of California Santa Cruz, USA)*
Guido Zuccon *(Queensland University of Technology, Australia)*

Local Arrangements:
Jean Joyce *(University of Massachusetts Amherst, USA)*
Kate Moruzzi *(University of Massachusetts Amherst, USA)*

Steering Committee Chair:
Oren Kurland *(Technion - Israel Institute of Technology, Israel)*

Steering Committee:
Leif Azzopardi *(University of Glasgow, UK)*
Peter Bruza *(Queensland University of Technology, Australia)*
Susan Dumais *(Microsoft Research, USA)*
Jaap Kamps, SIGIR representative *(University of Amsterdam, The Netherlands)*
Birger Larsen *(Aalborg University, Denmark)*
Donald Metzler *(Google, USA)*
Stefan Rueger, BCS-IRSG representative *(The Open University, UK)*
Keith van Rijsbergen, Honorary Chair *(University of Glasgow, UK)*

Additional reviewers:
Zhuyun Dai *(Carnegie Mellon University, USA)*
Kyle Yingkai Gao *(Carnegie Mellon University, USA)*
Parantapa Goswami *(University J. Fourier/Grenoble 1, France)*
Savvas Karagiannidis *(University College London, UK)*
Kleanthis Malialis *(University College London, UK)*
Stefania Marrara *(Universita degli Studi di Milano Bicocca, Italy)*
Daan Odijk *(University of Amsterdam, The Netherlands)*
Ridho Reinanda *(University of Amsterdam, The Netherlands)*
Anne Schuth *(University of Amsterdam, The Netherlands)*
Marc Sloan *(University College London, UK)*
Marco Viviani *(Universita degli Studi di Milano Bicocca, Italy)*
Weinan Zhang *(University College London, UK)*
Xiaoxue Zhao *(University College London, UK)*

ICTIR 2015 Sponsor & Supporters

Sponsor:

Organizer:

Supporters:

Silver

Bronze