# ICTIR 2013

International Conference on the Theory of Information Retrieval
29 September – 2 October 2013, Copenhagen, Denmark

Under the auspices of BCS-IRSG
In Cooperation with ACM SIGIR

ICTIR13 Proceedings

Oren Kurland, Donald Metzler, Christina Lioma, Birger Larsen, Peter Ingwersen (editors), Proceedings of the 4th International Conference on the Theory of Information Retrieval (ICTIR13), September 29 – October 2, 2013, Copenhagen, Denmark. ACM 2013, ISBN 978-1-4503-2107-5.

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

ACM COPYRIGHT NOTICE. Copyright © 2013 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of all or part 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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or permissions@acm.org.

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 the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

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 was 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.

ACM ISBN: 978-1-4503-2107-5

Cover photograph by Ty Stange, distributed by Wonderful Copenhagen.

# Table of Contents

**Preface** i
**ICTIR13 Organisation** iii
**ICTIR13 Program Committee** iv
**ICTIR13 Sponsors and Supporters** vi

**Tutorials**
Quantum Mechanics and Information Retrieval: From Theory to Application 1
Massimo Melucci (University of Padua)
Benjamin Piwowarski (Centre National de la Recherche Scientifique)

Statistical Significance Testing in Information Retrieval: Theory and Practice 2
Ben Carterette (University of Delaware)

Axiomatic Analysis and Optimization of Information Retrieval Models 3
ChengXiang Zhai (University of Illinois at Urbana-Champaign)
Hui Fang (University of Delaware)

IR Models: Foundations and Relationships 4
Thomas Roelleke (Queen Mary, University of London)

**Keynote Address**
Is there Space for Theory in Modern Commercial Search Engines? 5
Ricardo Baeza-Yates (Yahoo! Labs)

**Session 1: Relevance Feedback (Long Papers)**
A Theoretical Analysis of Pseudo-Relevance Feedback Models 6
Stéphane Clinchant (Xerox Research Centre Europe)
Éric Gaussier (Université Joseph Fourier (Grenoble 1))

Query-Performance Prediction Using Minimal Relevance Feedback 14
Olga Butman (Intel)
Anna Shtok (Technion - Israel Institute of Technology)
Oren Kurland (Technion - Israel Institute of Technology)
David Carmel (Yahoo! Labs)

**Session 2: Evaluation (Long Papers)**
Axiometrics: An Axiomatic Approach to Information Retrieval Effectiveness Metrics 22
Luca Busin (University of Udine)
Stefano Mizzaro (University of Udine)

On Using Fewer Topics in Information Retrieval Evaluations 30
Andrea Berto (University of Udine)
Stefano Mizzaro (University of Udine)
Stephen Robertson (University College London)

**Panel**
IR Challenges and Long-Range Opportunities 38
Peter Ingwersen (University of Copenhagen)

**Keynote Address**
Compositional Vector Semantics over Different Ground Fields 39
Dominic Widdows (Serendipity)

**Session 3: Recommender Systems (Long Papers)**
Enhanced Information Retrieval by Exploiting Recommender Techniques in Cluster-Based Link Analysis 40
Wei Li (Dublin City University)
Gareth Jones (Dublin City University)

Understanding Similarity Metrics in Neighbour-based Recommender Systems 48
Alejandro Bellogin (Centrum Wiskunde & Informatica)
Arjen de Vries (Centrum Wiskunde & Informatica)

**Session 4: Temporal & Thread Search (Long Papers)**
Information Retrieval with Time Series Query 56
Hyun Duk Kim (University of Illinois at Urbana-Champaign)
Danila Nikitin (University of Illinois at Urbana-Champaign)
Chengxiang Zhai (University of Illinois at Urbana-Champaign)
Malu Castellanos (Hewlett Packard Labs)
Meichun Hsu (Hewlett Packard Labs)

Exploiting Forum Thread Structures to Improve Thread Clustering 64
Kumaresh Pattabiraman (University of Illinois at Urbana-Champaign)
Parikshit Sondhi (University of Illinois at Urbana-Champaign)
Chengxiang Zhai (University of Illinois at Urbana-Champaign)

**Session 5: Context & Diversification (Short Papers)**
Tie-Breaker: A Novel Way of Combining Retrieval Signals 72
Wu Hao (University of Delaware)
Hui Fang (University of Delaware)

A Diagnostic Study of Search Result Diversification Methods 76
Wei Zheng (University of Delaware)
Hui Fang (University of Delaware)

Opinion-based User Profile Modeling for Contextual Suggestions 80
Peilin Yang (University of Delaware)
Hui Fang (University of Delaware)

**Keynote Address**
The Roots of the Theoretical Basis for Information Retrieval 84
Keith van Rijsbergen (University of Cambridge)

**Session 6: Ranking I (Long Papers)**
Modelling Score Distributions Without Actual Scores 85
Stephen Robertson (University College London)
Evangelos Kanoulas (Google Inc.)
Emine Yilmaz (University College London)

Revisiting Exhaustivity and Specificity Using Propositional Logic and Lattice Theory 93
Karam Abdulahhad (Université Joseph Fourier Grenoble)
Jean-Pierre Chevallet (Université Joseph Fourier Grenoble)
Catherine Berrut (Université Joseph Fourier Grenoble)

Efficient Nearest-Neighbor Search in the Probability Simplex 101
Kriste Krstovski (University of Massachusetts Amherst)
David A. Smith (Northeastern University)
Hanna M. Wallach (University of Massachusetts Amherst)
Andrew McGregor (University of Massachusetts Amherst)

**Session 7: Ranking II (Short Papers)**
A New Probabilistic Model for IR 109
Richard Connor (University of Strathclyde)
Robert Moss (University of Strathclyde)
Morgan Harvey (University of Strathclyde)

A Standard Document Score for Information Retrieval 113
Ronan Cummins (University of Greenwich)

Textual Similarity with a Bag-of-Embedded-Words Model 117
Stéphane Clinchant (Xerox Research Centre Europe)
Florent Perronnin (Xerox Research Centre Europe)

Statistical Translation Language Model for Twitter Search 121
Maryam Karimzadehgan (University of Illinois at Urbana-Champaign)
ChengXiang Zhai (University of Illinois at Urbana-Champaign)
Miles Efron (University of Illinois at Urbana-Champaign)

**Posters**
More Than Words: A Review of Planets, Stars and Sample Spaces 125
Emanuele Di Buccio (University of Padua)
Giorgio Maria Di Nunzio (University of Padua)

A Visual Analysis of the Effects of Assumptions of Classical Probabilistic Models 127
Emanuele Di Buccio (University of Padua)
Giorgio Maria Di Nunzio (University of Padua)

Information Retrieval for Temporal Bounding 129
Leon Derczynski (University of Sheffield)
Robert Gaizauskas (University of Sheffield)

Mathematical Specification and Logic Modelling in the Context of IR 131
Miguel Martinez-Alvarez (Queen Mary, University of London)
Marco Bonzanini (Queen Mary, University of London)
Thomas Roelleke (Queen Mary, University of London)

A Modification of LambdaMART to Handle Noisy Crowdsourced Assessments 133
Pavel Metrikov (Northeastern University)
Jie Wu (Northeastern University)
Jesse Anderton (Northeastern University)
Virgil Pavlu (Northeastern University)
Javed Aslam (Northeastern University)

**Author Index** 135

# Preface

These proceedings contain the refereed papers, posters and abstracts of keynotes, tutorials and panel discussion presented at the Fourth International Conference on the Theory of Information Retrieval (ICTIR13), held in Copenhagen, Denmark, during September 29 - October 2, 2013.

ICTIR is held biennially under the auspices of BCS-IRSG and attracts the latest advances in theoretical and formal Information Retrieval (IR) research. As of this year, ICTIR is also held in cooperation with ACM SIGIR. Previous editions of ICTIR were held in Budapest, Hungary (October 2007), Cambridge, UK (September 2009), and Bertinoro, Italy (September 2011). We are very happy to bring ICTIR13 to Denmark, a country with a long history in IR. Continuing the formal tradition of the previous conference editions but also looking into the future of the area, for the first time ICTIR includes tutorials, a panel discussion, as well as a strong international (as opposed to mainly European) presence in its organisation and participation by leading researchers from both academia and industry.

The 11 full papers, 7 short papers, and 5 posters presented at ICTIR13 were carefully reviewed and selected from 51 submissions. The acceptance rates were 29% for full papers, 48% for full & short papers, and 57% for posters. Each submission was subject to a double-blind reviewing process by at least three program committee members. The technical program covers a wide range of areas, including evaluation, ranking, relevance feedback, context, diversification, recommender systems and temporal and thread search. For the first time, the highest share of the work presented at ICTIR is contributed from North America (36%), followed by the UK (21%), and the rest of the world. The participation of five students (two from the UK, and the rest from Canada, USA and the Netherlands) was supported through student grants very kindly provided by Google and ICTIR 2011 surplus. For the first time the technical program includes three keynote talks, two of which are from industry: We thank Ricardo Baeza-Yates (Yahoo! Labs) for his talk on the role of theory in commercial search engines; Dominic Widdows (Serendipity) for his talk on compositional vector semantics; and Keith van Rijsbergen (University of Cambridge) for his talk on the roots of the theoretical basis for IR. We hope that these keynote talks, which cover a range of different core aspects of theoretical IR, will inspire both junior and senior researchers alike and drive further research in the area.

We thank the program committee of papers, posters, tutorials and best paper award for their time and high-quality discussions and feedback to all authors. We also thank Massimo Melucci for his initial work as program committee co-chair, and BSC-IRSG member Jun Wang for his efforts, ideas and support. We especially thank Casper Petersen, the local organisation chair & website designer, Toine Bogers, the local entertainment chair, and the student volunteers. Finally, we are deeply thankful to the following sponsors and supporters for contributing in making ICTIR13 a success:

- Sponsors: Google, Yahoo!, LARM Audio Research Archive, Spinque, Findwise, the Department of Computer Science, the Royal School of Library and Information Science and the Centre for Communication and Computing of the University of Copenhagen.

- Supporters: BCS-IRSG, ACM SIGIR, Wonderful Copenhagen and the editorial staff at ACM for their agreement and assistance in publishing the conference proceedings.

September 2013

General Chairs:
*Christina Lioma*
*Birger Larsen*
*Peter Ingwersen*

Honorary Chair:
*Keith van Rijsbergen*

Program Chairs:
*Oren Kurland*
*Donald Metzler*

# ICTIR13 Organisation

**General Co-Chairs:**

Christina Lioma
(University of Copenhagen, Denmark)

Birger Larsen
(University of Copenhagen, Denmark)

Peter Ingwersen
(University of Copenhagen, Denmark)

**Honorary Chair:**

Keith van Rijsbergen
(University of Cambridge, UK)

**Program Co-Chairs:**

Oren Kurland
(Technion --- Israel Institute of Technology, Israel)

Donald Metzler
(Google, USA)

**Best Paper Committee Chair:**

Bruce Croft
(University of Massachusetts Amherst, USA)

**Local Organisation Chair:**

Casper Petersen
(University of Copenhagen, Denmark)

**Local Entertainment Chair:**

Toine Bogers
(University of Copenhagen, Denmark)

**Webmaster**:

Casper Petersen
(University of Copenhagen, Denmark)

# ICTIR13 Program Committee

| | |
|---|---|
| James Allan | University of Massachusetts Amherst, USA |
| Giambattista Amati | Fondazione Ugo Bordoni, Italy |
| Jaime Arguello | University of North Carolina at Chapel Hill, USA |
| Javed Aslam | Northeastern University, USA |
| Leif Azzopardi | University of Glasgow, UK |
| Krisztian Balog | University of Stavanger, Norway |
| Michael Bendersky | Google, USA |
| Jiang Bian | Microsoft Research Asia, China |
| Roi Blanco | Yahoo! Labs, Spain |
| Peter Bruza | Queensland University of Technology, Australia |
| Jamie Callan | Carnegie Mellon University, USA |
| Claudio Carpineto | Fondazione Ugo Bordoni, Italy |
| Ben Carterette | University of Delaware, USA |
| Charles Clarke | University of Waterloo, Canada |
| Fabio Crestani | University of Lugano, Switzerland |
| Bruce Croft | University of Massachusetts Amherst, USA |
| Hui Fang | University of Delaware, USA |
| Ingo Frommholz | University of Bedfordshire, UK |
| Norbert Fuhr | University of Duisburg-Essen, Germany |
| Eric Gaussier | Université Joseph Fourier (Grenoble 1), France |
| Claudia Hauff | Delft University of Technology, Netherlands |
| Jimmy Huang | York University, Canada |
| Peter Ingwersen | University of Copenhagen, Denmark |
| Franciska de Jong | University of Twente, Netherlands |
| Gareth Jones | Dublin City University, Ireland |
| Joemon Jose | University of Glasgow, UK |
| Noriko Kando | National Institute of Informatics, Japan |
| Oren Kurland | Technion --- Israel Institute of Technology, Israel |
| Mounia Lalmas | Yahoo! Labs, Spain |
| Birger Larsen | University of Copenhagen, Denmark |
| Christina Lioma | University of Copenhagen, Denmark |
| Yuanhua Lv | Microsoft Research, USA |
| R. Manmatha | University of Massachusetts Amherst, USA |
| Donald Metzler | Google, USA |
| Alistair Moffat | University of Melbourne, Australia |
| Jian-Yun Nie | Université de Montréal, Canada |
| Iadh Ounis | University of Glasgow, UK |
| Gabriella Pasi | Università di Milano Bicocca, Italy |
| Virgil Pavlu | Northeastern University, USA |
| Benjamin Piwowarski | Centre National de la Recherche Scientifique, France |
| Keith van Rijsbergen | University of Cambridge, UK |
| Stephen Robertson | University College London, UK |
| Thomas Roelleke | Queen Mary University of London, UK |
| Rodrygo L. T. Santos | Universidade Federal de Minas Gerais, Brazil |
| Alan Smeaton | Dublin City University, Ireland |
| Ian Soboroff | National Institute of Standards and Technology, USA |
| Dawei Song | Robert Gordon University, UK |
| Kevyn Collins-Thompson | Microsoft Research, USA |
| Arjen de Vries | Centrum Wiskunde & Informatica, Netherlands |
| Jun Wang | University College London, UK |
| Lidan Wang | University of Maryland, USA |
| Dominic Widdows | Serendipity, USA |
| Xiaobing Xue | Twitter, USA |

| | |
|---|---|
| Yuexian Hou | Tianjin University, China |
| Dell Zhang | Birkbeck, University of London, UK |
| Guido Zuccon | Australian e-Health Research Centre, Australia |

**Posters Committee**

| | |
|---|---|
| Avi Arampatzis | Democritus University of Thrace, Greece |
| Roi Blanco | Yahoo! Labs, Spain |
| Ronan Cummins | University of Greenwich, UK |
| Giorgio Maria Di Nunzio | University of Padua, Italy |
| Gabriella Pasi | Università di Milano Bicocca, Italy |
| Virgil Pavlu | Northeastern University, USA |
| Guido Zuccon | Australian e-Health Research Centre, Australia |

**Tutorials Committee**

| | |
|---|---|
| Ram Akella | University of California, Berkeley & Santa Cruz, USA |
| Craig Macdonald | University of Glasgow, UK |
| Fabrizio Silvestri | Consiglio Nazionale delle Ricerche, Italy |
| Mark Smucker | University of Waterloo, Canada |
| Tao Qin | Microsoft Research Asia, China |

**Best Papers Committee**

| | |
|---|---|
| James Allan | University of Massachusetts Amherst, USA |
| Michael Bendersky | Google, USA |
| Bruce Croft | University of Massachusetts Amherst, USA |
| ChengXiang Zhai | University of Illinois at Urbana-Champaign, USA |

# ICTIR13 Sponsors and Supporters

## Sponsors

## Supporters