

September 14–17, 2018  
Tianjin, China



Association for  
Computing Machinery

*Advancing Computing as a Science & Profession*



# ICTIR'18

Proceedings of the 2018 ACM SIGIR International Conference on  
Theory of Information Retrieval

*Sponsored by:*

**ACM SIGIR**

*Supported by:*

***Sogou, Tianjin University, Microsoft Research,  
Chinese Information Processing Society of China,  
and BCS Information Retrieval Specialist Group***



**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 © 2018 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](mailto: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](http://www.copyright.com).

**ISBN:** 978-1-4503-5656-5

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](mailto:acmhelp@acm.org)  
Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA

# General Chairs' Welcome

Welcome to ICTIR 2018, the 8th International Conference on the Theory of Information

Retrieval and the 4th conference with that name to be fully sponsored by the ACM Special Interest Group on Information Retrieval (SIGIR).

We thank everyone who helped us organize the conference: Long paper chairs Peter Bruza and Massimo Melucci, short paper chairs Fabrizio Sebastiani and Grace Hui Yang, workshop chair Emanuele Di Buccio, tutorial chair Haiming Liu, sponsorship/industry track chairs Zhicheng Dou and Udo Kruschwitz, publicity chairs Jiafeng Guo and Ingo Frommholz, local organization chairs Peng Zhang, Bo Wang and Yuexian Hou, proceedings chairs Sagar Uprety and Qiuchi Li, webmaster Cheng Zhang, and the members of the program committee who worked hard to review papers in a short time-frame and provide feedback for authors.

We also thank our sponsors and supporters: ACM SIGIR, Sougou, Tianjin University, Chinese Information Processing Society of China, Microsoft Research Asia and British Computer Society's Information Retrieval Specialists Group.

Finally, we thank the authors for providing superb technical content, the workshop organizers, the tutorial lecturers and our keynote speakers Ming Zhou and Yiming Yang.

We look forward to welcoming you in Tianjin, China, and hope you enjoy your stay.

**Dawei Song, Tie-Yan Liu & Le Sun**

*ICTIR'18 General Chairs*

*Beijing Institute of Technology and The Open University,  
Microsoft Research Asia & Chinese Information Processing  
Society of China*

# Programme Chairs' Welcome

ACM SIGIR ICTIR2018 conference provides a forum for the presentation and discussion of research related to the fundamental aspects of Information Retrieval (IR), including human search processes, search and recommendation, learning and optimization, language and representations, sensory information, artificial intelligence, ethics and responsibility in information retrieval.

The call for papers attracted submissions from all over the world. The program committee reviewed 70 submissions – 47 long and papers and 23 short papers. Of these 19 long papers and 16 short papers were accepted for presentation at the conference. We also accepted two tutorial proposals and a workshop proposal. We sincerely hope that these papers, tutorials and workshop will help lay down and promote foundational ideas in information retrieval research.

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 institutions around the world.

**Peter Bruza & Massimo Melucci**

*ICTIR'18 Program Chairs (Long Papers)*

*Queensland University of Technology*

*& University of Padua*

**Fabrizio Sebastiani & Grace Hui Yang**

*ICTIR'18 Program Chairs (Short Papers)*

*Consiglio Nazionale delle Ricerche*

*& Georgetown University*

# Table of Contents

ICTIR'18 Organization .....	viii
-----------------------------	------

ICTIR'18 Sponsor & Supporters .....	xi
-------------------------------------	----

## Tutorials

• <b>Unbiased Learning to Rank: Theory and Practice</b> .....	1
Qingyao Ai ( <i>University of Massachusetts, Amherst</i> ), Jiaxin Mao, Yiqun Liu ( <i>Tsinghua University</i> ), W. Bruce Croft ( <i>University of Massachusetts, Amherst</i> )	
• <b>User Interactions with Search Systems</b> .....	3
Ying-Hsang Liu ( <i>Charles Sturt University</i> ), Chang Liu ( <i>Peking University</i> ), Ralf Bierig ( <i>Maynooth University</i> )	

## Workshops

• <b>The 8th Symposium on Future Directions in Information Access</b> .....	5
Haiming Liu, Ingo Frommholz ( <i>University of Bedfordshire</i> ), Ingo Schmitt ( <i>BTU Cottbus-Senftenberg</i> ), Dawei Song ( <i>Beijing Institute of Technology</i> )	

## Keynotes

• <b>What Will Search Engines be Changed by NLP Advancements</b> .....	7
Ming Zhou ( <i>Microsoft Research Asia</i> )	
• <b>Large-scale Machine Learning over Graphs</b> .....	9
Yiming Yang ( <i>Carnegie Mellon University</i> )	

## Long Paper Session 1: Collaborative Ranking/Filtering and Recommendations

• <b>A Word is Worth a Thousand Ratings: Augmenting Ratings using Reviews for Collaborative Filtering</b> .....	11
Oren Sar Shalom ( <i>Intuit</i> ), Guy Uziel ( <i>IBM Research</i> ), Alexandros Karatzoglou ( <i>Telefonica Research</i> ), Amir Kantor ( <i>IBM Research</i> )	
• <b>A Collaborative Ranking Model with Multiple Location-based Similarities for Venue Suggestion</b> .....	19
Mohammad Aliannejadi ( <i>Università della Svizzera italiana (USI)</i> ), Dimitrios Rafailidis ( <i>University of Mons</i> ), Fabio Crestani ( <i>Università della Svizzera italiana (USI)</i> )	
• <b>Attentive Contextual Denoising Autoencoder for Recommendation</b> .....	27
Yogesh Jhamb, Travis Ebesu, Yi Fang ( <i>Santa Clara University</i> )	

## Long Paper Session 2: Queries and Browsing

• <b>An Extended Query Performance Prediction Framework Utilizing Passage-Level Information</b> .	35
Haggai Roitman ( <i>IBM Research</i> )	
• <b>On the Equilibrium of Query Reformulation and Document Retrieval</b> .....	43
Shihao Zou, Guanyu Tao, Jun Wang ( <i>University College London</i> ), Weinan Zhang ( <i>Shanghai Jiao Tong University</i> ), Dell Zhang ( <i>Birbeck, University of London</i> )	
• <b>Generating Synthetic Data for Neural Keyword-to-Question Models</b> .....	51
Heng Ding ( <i>Wuhan University</i> ), Krisztian Balog ( <i>University of Stavanger</i> )	

## Long Paper Session 3: Neural IR

• <b>Empirical Study of Multi-level Convolution Models for IR Based on Representations and Interactions</b> .....	59
Yifan Nie ( <i>Université de Montréal</i> ), Yanling Li ( <i>Inner Mongolia Normal University</i> ), Jian-Yun Nie ( <i>Université de Montréal</i> )	

- **Understanding the Representational Power of Neural Retrieval Models Using NLP Tasks** ..... 67  
Daniel Cohen, Brendan O'Connor, W. Bruce Croft (*University of Massachusetts, Amherst*)
- **Product Question Intent Detection using Indicative Clause Attention and Adversarial Learning** ..... 75  
Qian Yu, Wai Lam (*Chinese University of Hong Kong*)

### Long Paper Session 4: Efficiency

- **MQGrad: Reinforcement Learning of Gradient Quantization in Parameter Server** ..... 83  
Guoxin Cui, Jun Xu, Wei Zeng, Yanyan Lan, Jiafeng Guo, Xueqi Cheng (*University of Chinese Academy of Sciences & Chinese Academy of Sciences*)
- **Measuring the Effectiveness of Selective Search Index Partitions without Supervision** ..... 91  
Yubin Kim, Jamie Callan (*Carnegie Mellon University*)
- **Beyond Greedy Search: Pruned Exhaustive Search for Diversified Result Ranking** ..... 99  
Yingying Wu (*The University of Texas at Austin*), Yiqun Liu, Fei Chen, Min Zhang, Shaoping Ma (*Tsinghua University*)
- **A Vertical PRF Architecture for Microblog Search** ..... 107  
Flávio Martins, João Magalhães (*Universidade NOVA de Lisboa*), Jamie Callan (*Carnegie Mellon University*)

### Long Paper Session 5: Theory of Evaluation

- **Using PageRank for Characterizing Topic Quality in LDA** ..... 115  
Sujatha Das Gollapalli, Xiao-li Li (*A\*STAR*)
- **A Formal Account of Effectiveness Evaluation and Ranking Fusion** ..... 123  
Enrique Amigó, Fernando Giner (*UNED NLP & IR Group*), Stefano Mizzaro (*University of Udine*), Damiano Spina (*RMIT University*)
- **The Broad View of Task Type Using Path Analysis** ..... 131  
Matthew Mitsui, Chirag Shah (*Rutgers University*)

### Long Paper Session 6: Theory of Models

- **Pseudo Descriptions for Meta-Data Retrieval** ..... 139  
Tim Gollub, Erdan Genc (*Bauhaus-Universität Weimar*), Nedim Lipka (*Adobe Systems*), Benno Stein (*Bauhaus-Universität Weimar*)
- **On the Theory of Weak Supervision for Information Retrieval** ..... 147  
Hamed Zamani, W. Bruce Croft (*University of Massachusetts, Amherst*)
- **Entity Set Expansion from Twitter** ..... 155  
He Zhao, Chong Feng (*Beijing Institute of Technology*), Zhunchen Luo, Changhai Tian (*PLA Academy of Military Science*)

### Short Paper Session 1: AI and IR

- **Entire Information Attentive GRU for Text Representation** ..... 163  
Guoxiu He, Wei Lu (*Wuhan University*)
- **Towards Word Embeddings for Improved Duplicate Bug Report Retrieval in Software Repositories** ..... 167  
Amar Budhiraja, Kartik Dutta, Manish Shrivastava, Raghu Reddy (*IIT-Hyderabad*)
- **Levenshtein in Blocks World: String Matching via AI Planning** ..... 171  
Xing Tan, Jimmy Xiangji Huang (*York University*)
- **Multi Page Search with Reinforcement Learning to Rank** ..... 175  
Wei Zeng, Jun Xu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng (*University of Chinese Academy of Sciences & Chinese Academy of Sciences*)
- **Multitask Learning for Query Segmentation in Job Search** ..... 179  
Bahar Salehi, Fei Liu, Timothy Baldwin (*University of Melbourne*), Wilson Wong (*SEEK Ltd.*)

## Short Paper Session 2: Quantum-inspired IR Models

- **Quantum Language Model-based Query Expansion** ..... 183  
Qiuchi Li, Massimo Melucci, Prayag Tiwari (*University of Padova*)
- **Binary Classification Model Inspired from Quantum Detection Theory** ..... 187  
Emanuele Di Buccio, Qiuchi Li, Massimo Melucci, Prayag Tiwari (*University of Padova*)
- **Investigating Order Effects in Multidimensional Relevance Judgment using Query Logs**..... 191  
Sagar Uprety, Dawei Song (*Open University*)

## Short Paper Session 3: Fusion and Evaluation

- **Enhanced Performance Prediction of Fusion-based Retrieval**..... 195  
Haggai Roitman (*IBM Research*)
- **Topic Set Size Design for Paired and Unpaired Data** ..... 199  
Tetsuya Sakai (*Waseda University*)
- **Utilizing Pseudo-Relevance Feedback in Fusion-based Retrieval** ..... 203  
Haggai Roitman (*IBM Research*)
- **StatBM25: An Aggregative and Statistical Approach for Document Ranking** ..... 207  
Xing Tan (*York University*), Fanghong Jiang (*Central China Normal University*), Jimmy Xiangji Huang (*York University*)

## Short Paper Session 4: Entity Retrieval and Question Answering

- **Entity Retrieval in the Knowledge Graph with Hierarchical Entity Type and Content**..... 211  
Xinshi Lin, Wai Lam, Kwun Ping Lai (*Chinese University of Hong Kong*)
- **An Adaptive Recommender System for Computational Serendipity**..... 215  
Xi Niu (*University of North Carolina at Charlotte*)
- **Classifying Community QA Questions That Contain an Image** ..... 219  
Kenta Tamaki (*Waseda University*), Riku Togashi (*Yahoo Japan Corporation*), Sosuke Kato (*Waseda University*),  
Sumio Fujita (*Yahoo Japan Corporation*), Hideyuki Maeda (*CyberAgent, Inc.*), Tetsuya Sakai (*Waseda University*)
- **Exploring Diversification In Non-factoid Question Answering** ..... 223  
Lakshmi Vikraman, W. Bruce Croft, Brendan O'Connor (*University of Massachusetts*)

- **Author Index** ..... 227

# ICTIR 2018 Conference Organization

**General Chairs:** Dawei Song (*Beijing Institute of Technology, China and The Open University, UK*)  
Tie-Yan Liu (*Microsoft Research Asia, China*)  
Le Sun (*Chinese Information Processing Society of China, China*)

**Program Chairs (Long Papers):** Peter Bruza (*Queensland University of Technology, Australia*)  
Massimo Melucci (*University of Padua, Italy*)

**Program Chairs (Short Papers):** Fabrizio Sebastiani (*Consiglio Nazionale delle Ricerche, Italy*)  
Grace Hui Yang (*Georgetown University, USA*)

**Workshop Chair:** Emanuele Di Buccio (*University of Padua, Italy*)

**Tutorial Chair:** Haiming Liu (*University of Bedfordshire, UK*)

**Sponsorship/Industry** Zhicheng Dou (*Renmin University of China, China*)

**Track Chairs:** Udo Kruschwitz (*University of Essex, UK*)

**Publicity Chairs:** Jiafeng Guo (*Chinese Academy of Sciences, China*)  
Ingo Frommholz (*University of Bedfordshire, UK*)

**Local Organization Chairs:** Peng Zhang (*Tianjin University, China*)  
Bo Wang (*Tianjin University, China*)  
Yuexian Hou (*Tianjin University, China*)

**Proceedings Chairs:** Sagar Uprety (*The Open University, UK*)  
Qiuchi Li (*University of Padua, Italy*)

**Webmaster:** Cheng Zhang (*Tianjin University, China*)

**Program Committee:** Eugene Agichtein (*Emory University, USA*)  
Giambattista Amati (*Fondazione Ugo Bordoni, Italy*)  
Enrique Amigó (*UNED, Spain*)  
Jaime Arguello (*The University of North Carolina at Chapel Hill, USA*)  
Leif Azzopardi (*University of Strathclyde, UK*)  
Philip Bachman (*Microsoft, USA*)  
Alberto Barrón-Cedeño (*Qatar Computing Research Institute, Qatar*)  
Gaurav Baruah (*Klick Inc., Canada*)  
Nick Belkin (*Rutgers University, USA*)  
Pavel Braslavski (*Ural Federal University, Russia*)  
Jamie Callan (*Carnegie Mellon University, USA*)  
Claudio Carpineto (*Fondazione Ugo Bordoni, Italy*)  
Ingemar Cox (*University College London, UK*)  
Fabio Crestani (*University of Lugano (USI), Italy*)  
Bruce Croft (*University of Massachusetts Amherst, USA*)  
Ronan Cummins (*University of Cambridge, UK*)



**Program Committee** Arjen P. de Vries (*Radboud University, Netherlands*)  
**(continued):** Emanuele Di Buccio (*University of Padua, Italy*)  
Tamer El Sayed (*Qatar University, Qatar*)  
Nicola Ferro (*University of Padua, Italy*)  
Norbert Fuhr (*University of Duisburg-Essen, Germany*)  
Eric Gaussier (*LIG-UJF, France*)  
Shlomo Geva (*Queensland University of Technology, Australia*)  
Ben He (*University of Chinese Academy of Sciences, China*)  
Jiyin He (*Signal Media Ltd., UK*)  
Jimmy Huang (*University of York, Canada*)  
Jaap kamps (*University of Amsterdam, Netherlands*)  
Evangelos Kanoulas (*University of Amsterdam, Netherlands*)  
Diane Kelly (*University of Tennessee, USA*)  
Bevan Koopman (*CSIRO, Australia*)  
Oren Kurland (*Technion, Israel Institute of Technology, Israel*)  
Wai Lam (*The Chinese University of Hong Kong, Hong Kong*)  
Cheng Li (*Google, USA*)  
Yiqun Liu (*Tsinghua University, China*)  
Joao Magalhaes (*Universidade NOVA de Lisboa, Portugal*)  
Bhaskar Mitra (*Microsoft, USA*)  
Boughanem Mohand (*IRIT University Paul Sabatier Toulouse, France*)  
Jayanta Mukherjee (*IIT Kharagpur, India*)  
Jian-Yun Nie (*University of Montreal, Italy*)  
Benjamin Piwowarski (*CNRS / University Pierre et Marie Curie, France*)  
Salvatore Romeo (*Qatar Computing Research Institute, Qatar*)  
Stefan Rueger (*Knowledge Media Institute, UK*)  
Rodrygo Santos (*Universidade Federal de Minas Gerais, Brazil*)  
Alessandro Sordoni (*Microsoft, Canada*)  
Laure Soulier (*Sorbonne Universités UPMC-LIP6, France*)  
Lynda Tamine (*IRIT, France*)  
Paul Thomas (*Microsoft, USA*)  
Gabriele Tolomei (*University of Padua, Italy*)  
Andrew Trotman (*University of Otago, New Zealand*)  
Theodora Tsikrika (*Information Technologies Institute, CERTH, Greece*)  
Ellen M. Voorhees (*NIST, USA*)  
Jiyin He (*Signal Media Ltd., UK*)  
Shlomo Geva (*Queensland University of Technology, Australia*)  
Yue Wang (*University of Michigan, USA*)  
Xuanhui Wang (*Google, USA*)  
Hao Wu (*University of Delaware, USA*)  
Liu Yang (*University of Massachusetts Amherst, USA*)  
Yinan Zhang (*University of Illinois at Urbana-Champaign, USA*)  
Yi Zhang (*University of California, Santa Cruz, USA*)  
Guido Zuccon (*Queensland University of Technology, Australia*)

**Additional reviewers:** Mohammad Aliannejadi  
Artem Babenko  
Seyed Ali Bahrainian  
Dan Cohen  
Maram Hasanain  
Mucahid Kutlu

Monica Landoni  
Flavio Martins  
André Mourão  
Chen Qu  
Reem Suwaileh

# ICTIR 2018 Sponsors & Supporters

## Sponsor



## Supporters



## In Collaboration With

