Contents

Social Context and News

  Minh-Tien Nguyen and Minh-Le Nguyen

A Graph-Based Approach to Topic Clustering for Online Comments to News ......................................................... 15
  Ahmet Aker, Emina Kurtic, A.R. Balamurali, Monica Paramita,
  Emma Barker, Mark Hepple, and Rob Gaizauskas

Leveraging Semantic Annotations to Link Wikipedia and News Archives ................................................................. 30
  Arunav Mishra and Klaus Berberich

Machine Learning

Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction ................................. 45
  Weinan Zhang, Tianming Du, and Jun Wang

Supervised Local Contexts Aggregation for Effective Session Search ................................................................. 58
  Zhiwei Zhang, Jingang Wang, Tao Wu, Pengjie Ren, Zhumin Chen,
  and Luo Si

An Empirical Study of Skip-Gram Features and Regularization for Learning on Sentiment Analysis ......................... 72
  Cheng Li, Bingyu Wang, Virgil Pavlu, and Javed A. Aslam

Multi-task Representation Learning for Demographic Prediction ........................................................................ 88
  Pengfei Wang, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng

Large-Scale Kernel-Based Language Learning Through the Ensemble Nyström Methods ........................................ 100
  Danilo Croce and Roberto Basili

Question Answering

Beyond Factoid QA: Effective Methods for Non-factoid Answer Sentence Retrieval ............................................... 115
  Liu Yang, Qingyao Ai, Damiano Spina, Ruey-Cheng Chen, Liang Pang,
  W. Bruce Croft, Jiafeng Guo, and Falk Scholer
Supporting Human Answers for Advice-Seeking Questions in CQA Sites . . . 129  
Liora Braunstein, Oren Kurland, David Carmel, Idan Szpektor, 
and Anna Shtok

Ranking

Does Selective Search Benefit from WAND Optimization? . . . . . . . . . . . . . 145  
Yubin Kim, Jamie Callan, J. Shane Culpepper, and Alistair Moffat
Efficient AUC Optimization for Information Ranking Applications . . . . . 159  
Sean J. Welleck
Modeling User Interests for Zero-Query Ranking . . . . . . . . . . . . . . . . . . . . 171  
Liu Yang, Qi Guo, Yang Song, Sha Meng, Milad Shokouhi, 
Kieran McDonald, and W. Bruce Croft

Evaluation Methodology

Adaptive Effort for Search Evaluation Metrics . . . . . . . . . . . . . . . . . . . . . . 187  
Jiepu Jiang and James Allan
Evaluating Memory Efficiency and Robustness of Word Embeddings . . . . 200  
Johannes Jurgovsky, Michael Granitzer, and Christin Seifert
Characterizing Relevance on Mobile and Desktop . . . . . . . . . . . . . . . . . . . 212  
Manisha Verma and Emine Yilmaz

Probabilistic Modelling

Probabilistic Local Expert Retrieval . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 227  
Wen Li, Carsten Eickhoff, and Arjen P. de Vries
Probabilistic Topic Modelling with Semantic Graph . . . . . . . . . . . . . . . . . 240  
Long Chen, Joemon M. Jose, Haitao Yu, Fajie Yuan, and Huaizhi Zhang
Estimating Probability Density of Content Types for Promoting Medical 
Records Search . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 252  
Yun He, Qinmin Hu, Yang Song, and Liang He

Evaluation Issues

The Curious Incidence of Bias Corrections in the Pool . . . . . . . . . . . . . . 267  
Aldo Lipani, Mihai Lupu, and Allan Hanbury
Understandability Biased Evaluation for Information Retrieval . . . . . . . . . 280  
Guido Zuccon
The Relationship Between User Perception and User Behaviour in Interactive Information Retrieval Evaluation
  
  Mengdie Zhuang, Elaine G. Toms, and Gianluca Demartini

Multimedia

Using Query Performance Predictors to Improve Spoken Queries
  
  Jaime Arguello, Sandeep Avula, and Fernando Diaz

Fusing Web and Audio Predictors to Localize the Origin of Music Pieces for Geospatial Retrieval
  
  Markus Schedl and Fang Zhou

Key Estimation in Electronic Dance Music
  
  Ángel Faraldo, Emilia Gómez, Sergi Jordà, and Perfecto Herrera

Summarization

Evaluating Text Summarization Systems with a Fair Baseline from Multiple Reference Summaries
  
  Fahmida Hamid, David Haraburda, and Paul Tarau

Multi-document Summarization Based on Atomic Semantic Events and Their Temporal Relationships
  
  Yllias Chali and Mohsin Uddin

Tweet Stream Summarization for Online Reputation Management
  
  Jorge Carrillo-de-Albornoz, Enrique Amigó, Laura Plaza, and Julio Gonzalo

Reproducibility

Who Wrote the Web? Revisiting Influential Author Identification Research Applicable to Information Retrieval
  
  Martin Potthast, Sarah Braun, Tolga Buz, Fabian Duffhauß, Florian Friedrich, Jörg Marvin Gülzow, Jakob Köhler, Winfried Lötzsch, Fabian Müller, Maike Elisa Müller, Robert Paßmann, Bernhard Reinke, Lucas Rettenmeier, Thomas Rometsch, Timo Sommer, Michael Träger, Sebastian Wilhelm, Benno Stein, Efstathios Stamatatos, and Matthias Hagen

Toward Reproducible Baselines: The Open-Source IR Reproducibility Challenge
  
  Jimmy Lin, Matt Crane, Andrew Trotman, Jamie Callan, Ishan Chattopadhyaya, John Foley, Grant Ingersoll, Craig Macdonald, and Sebastiano Vigna
Experiments in Newswire Summarisation ........................................ 421  
Stuart Mackie, Richard McCreadie, Craig Macdonald, and Iadh Ounis

On the Reproducibility of the TAGME Entity Linking System ............ 436  
Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg

Twitter

Correlation Analysis of Reader’s Demographics and Tweet Credibility  
Perception ................................................................. 453  
Shafiza Mohd Shariff, Mark Sanderson, and Xiuzhen Zhang

Topic-Specific Stylistic Variations for Opinion Retrieval on Twitter ..... 466  
Anastasia Giachanou, Morgan Harvey, and Fabio Crestani

Inferring Implicit Topical Interests on Twitter .............................. 479  
Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri,  
and Mohsen Kahani

Topics in Tweets: A User Study of Topic Coherence Metrics for Twitter  
Data ................................................................. 492  
Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel

Retrieval Models

Supporting Scholarly Search with Keyqueries ................................ 507  
Matthias Hagen, Anna Beyer, Tim Gollub, Kristof Komlossy,  
and Benno Stein

Pseudo-Query Reformulation .................................................. 521  
Fernando Diaz

VODUM: A Topic Model Unifying Viewpoint, Topic and Opinion  
Discovery ................................................................. 533  
Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem,  
and Karen Pinel-Sauvagnat

Applications

Harvesting Training Images for Fine-Grained Object Categories Using  
Visual Descriptions .................................................. 549  
Josiah Wang, Katja Markert, and Mark Everingham

Do Your Social Profiles Reveal What Languages You Speak? Language  
Inference from Social Media Profiles .................................. 561  
Yu Xu, M. Rami Ghorab, Zhongqing Wang, Dong Zhou,  
and Séamus Lawless
Collaborative Filtering

Implicit Look-Alike Modelling in Display Ads – Transfer Collaborative Filtering to CTR Estimation
Weinan Zhang, Lingxi Chen, and Jun Wang

Efficient Pseudo-Relevance Feedback Methods for Collaborative Filtering Recommendation
Daniel Valcarce, Javier Parapar, and Álvaro Barreiro

Language Models for Collaborative Filtering Neighbourhoods
Daniel Valcarce, Javier Parapar, and Álvaro Barreiro

Adaptive Collaborative Filtering with Extended Kalman Filters and Multi-armed Bandits
Jean-Michel Renders

Short Papers

A Business Zone Recommender System Based on Facebook and Urban Planning Data.
Jovian Lin, Richard J. Oentaryo, Ee-Peng Lim, Casey Vu, Adrian Vu, Agus T. Kwee, and Philips K. Prasetyo

On the Evaluation of Tweet Timeline Generation Task
Walid Magdy, Tamer Elsayed, and Maram Hasanain

Finding Relevant Relations in Relevant Documents
Michael Schuhmacher, Benjamin Roth, Simone Paolo Ponzetto, and Laura Dietz

Probabilistic Multileave Gradient Descent
Harrie Oosterhuis, Anne Schuth, and Maarten de Rijke

Real-World Expertise Retrieval: The Information Seeking Behaviour of Recruitment Professionals.
Tony Russell-Rose and Jon Chamberlain

Compressing and Decoding Term Statistics Time Series.
Jinfeng Rao, Xing Niu, and Jimmy Lin
Feedback or Research: Separating Pre-purchase from Post-purchase Consumer Reviews ........................................... 682
   Mehedi Hasan, Alexander Kotov, Aravind Mohan, Shiyong Lu,
   and Paul M. Stieg

Inferring the Socioeconomic Status of Social Media Users Based on
Behaviour and Language ........................................... 689
   Vasileios Lampos, Nikolaos Aletras, Jens K. Geyti, Bin Zou,
   and Ingemar J. Cox

Two Scrolls or One Click: A Cost Model for Browsing Search Results .... 696
   Leif Azzopardi and Guido Zuccon

Determining the Optimal Session Interval for Transaction Log Analysis
of an Online Library Catalogue ................................... 703
   Simon Wakeling and Paul Clough

A Comparison of Deep Learning Based Query Expansion
with Pseudo-Relevance Feedback and Mutual Information ............ 709
   Mohannad ALMasri, Catherine Berrut, and Jean-Pierre Chevallet

A Full-Text Learning to Rank Dataset for Medical Information Retrieval .... 716
   Vera Boteva, Demian Gholipour, Artem Sokolov, and Stefan Riezler

Multi-label, Multi-class Classification Using Polylingual Embeddings ...... 723
   Georgios Balikas and Massih-Reza Amini

Learning Word Embeddings from Wikipedia for Content-Based
Recommender Systems .............................................. 729
   Cataldo Musto, Giovanni Semeraro, Marco de Gemmis,
   and Pasquale Lops

Tracking Interactions Across Business News, Social Media, and Stock
Fluctuations .......................................................... 735
   Ossi Karkulahti, Lidia Pivovarova, Mian Du, Jussi Kangasharju,
   and Roman Yangarber

Subtopic Mining Based on Three-Level Hierarchical Search Intentions ....... 741
   Se-Jong Kim, Jaehun Shin, and Jong-Hyeok Lee

Cold Start Cumulative Citation Recommendation for Knowledge Base
Acceleration .......................................................... 748
   Jingang Wang, Jingtian Jiang, Lejian Liao, Dandan Song,
   Zhiwei Zhang, and Chin-Yew Lin

Cross Domain User Engagement Evaluation ................................ 754
   Ali Montazeralghaem, Hamed Zamani, and Azadeh Shakery
An Empirical Comparison of Term Association and Knowledge Graphs for Query Expansion .................................................. 761
Saeid Balaneshinkordan and Alexander Kotov

Deep Learning to Predict Patient Future Diseases from the Electronic Health Records .................................................... 768
Riccardo Miotto, Li Li, and Joel T. Dudley

Improving Document Ranking for Long Queries with Nested Query Segmentation .......................................................... 775
Rishiraj Saha Roy, Anusha Suresh, Niloy Ganguly, and Monojit Choudhury

Sketching Techniques for Very Large Matrix Factorization ................................................................. 782
Raghavendran Balu, Teddy Furon, and Laurent Amsaleg

Diversifying Search Results Using Time: An Information Retrieval Method for Historians ................................................... 789
Dhruv Gupta and Klaus Berberich

On Cross-Script Information Retrieval ................................................................. 796
Nada Naji and James Allan

LeX: A Learning Approach for Local Expert Discovery on Twitter ................................................................. 803
Wei Niu, Zhijiao Liu, and James Caverlee

Clickbait Detection ................................................................. 810
Martin Potthast, Sebastian Köpsel, Benno Stein, and Matthias Hagen

Informativeness for Adhoc IR Evaluation: A Measure that Prevents Assessing Individual Documents ................................................... 818
Romain Deveaud, Véronique Moriceau, Josiane Mothe, and Eric SanJuan

What Multimedia Sentiment Analysis Says About City Liveability ................................................................. 824
Joost Boonzajer Flaes, Stevan Rudinac, and Marcel Worring

Demos

Scenemash: Multimodal Route Summarization for City Exploration ................................................................. 833
Jorrit van den Berg, Stevan Rudinac, and Marcel Worring

Exactus Like: Plagiarism Detection in Scientific Texts ................................................................. 837
Ilya Sochenkov, Denis Zubarev, Ilya Tikhomirov, Ivan Smirnov, Artem Shelmanov, Roman Suvorov, and Gennady Osipov

Jitter Search: A News-Based Real-Time Twitter Search Interface ................................................................. 841
Flávio Martins, João Magalhães, and Jamie Callan
### Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TimeMachine: Entity-Centric Search and Visualization of News Archives</td>
<td>845</td>
</tr>
<tr>
<td><em>Pedro Saleiro, Jorge Teixeira, Carlos Soares, and Eugénio Oliveira</em></td>
<td></td>
</tr>
<tr>
<td>OPMES: A Similarity Search Engine for Mathematical Content</td>
<td>849</td>
</tr>
<tr>
<td><em>Wei Zhong and Hui Fang</em></td>
<td></td>
</tr>
<tr>
<td>SHAMUS: UFAL Search and Hyperlinking Multimedia System</td>
<td>853</td>
</tr>
<tr>
<td><em>Petra Galuščáková, Shadi Saleh, and Pavel Pecina</em></td>
<td></td>
</tr>
<tr>
<td><strong>Industry Day</strong></td>
<td></td>
</tr>
<tr>
<td>Industry Day Overview</td>
<td>859</td>
</tr>
<tr>
<td><em>Omar Alonso and Pavel Serdyukov</em></td>
<td></td>
</tr>
<tr>
<td><strong>Workshops</strong></td>
<td></td>
</tr>
<tr>
<td>Bibliometric-Enhanced Information Retrieval: 3rd International BIR</td>
<td>865</td>
</tr>
<tr>
<td>Workshop</td>
<td></td>
</tr>
<tr>
<td><em>Philipp Mayr, Ingo Frommholz, and Guillaume Cabanac</em></td>
<td></td>
</tr>
<tr>
<td>MultiLingMine 2016: Modeling, Learning and Mining for Cross/Multilinguality</td>
<td>869</td>
</tr>
<tr>
<td>*Dino Ienco, Mathieu Roche, Salvatore Romeo, Paolo Rosso,</td>
<td></td>
</tr>
<tr>
<td>and Andrea Tagarelli</td>
<td></td>
</tr>
<tr>
<td>Proactive Information Retrieval: Anticipating Users’ Information Need</td>
<td>874</td>
</tr>
<tr>
<td><em>Sumit Bhatia, Debapriyo Majumdar, and Nitish Aggarwal</em></td>
<td></td>
</tr>
<tr>
<td>First International Workshop on Recent Trends in News Information</td>
<td>878</td>
</tr>
<tr>
<td>Retrieval (NewsIR’16)</td>
<td></td>
</tr>
<tr>
<td>*Miguel Martinez-Alvarez, Udo Kruschwitz, Gabriella Kazai,</td>
<td></td>
</tr>
<tr>
<td>Frank Hopfgartner, David Corney, Ricardo Campos, and Dyaa Albakour</td>
<td></td>
</tr>
<tr>
<td><strong>Tutorials</strong></td>
<td></td>
</tr>
<tr>
<td>Collaborative Information Retrieval: Concepts, Models and Evaluation</td>
<td>885</td>
</tr>
<tr>
<td><em>Lynda Tamine and Laure Soulier</em></td>
<td></td>
</tr>
<tr>
<td>Group Recommender Systems: State of the Art, Emerging Aspects and</td>
<td>889</td>
</tr>
<tr>
<td>Techniques, and Research Challenges</td>
<td></td>
</tr>
<tr>
<td><em>Ludovico Boratto</em></td>
<td></td>
</tr>
<tr>
<td>Living Labs for Online Evaluation: From Theory to Practice</td>
<td>893</td>
</tr>
<tr>
<td><em>Anne Schuth and Krisztian Balog</em></td>
<td></td>
</tr>
</tbody>
</table>
Real-Time Bidding Based Display Advertising: Mechanisms
and Algorithms ............................................. 897

*Jun Wang, Shuai Yuan, and Weinan Zhang*

**Author Index** ...................................................... 903
Advances in Information Retrieval
Wolf, I.; Crestani, F.; Moens, M.-F.; Mothe, J.; Silvestri, F.; Di Nunzio, G.M.; Hauff, C.; Silvello, G. (Eds.)
2016, XXXIX, 906 p. 124 illus. in color., Softcover
ISBN: 978-3-319-30670-4