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, Ruy-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

309

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

335

Angel 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

351

Yllias Chali and Mohsin Uddin

Multi-document Summarization Based on Atomic Semantic Events and Their Temporal Relationships

Jorge Carrillo-de-Albornoz, Enrique Amigó, Laura Plaza, and Julio Gonzalo

378

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

393

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

408
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
Retrieving Hierarchical Syllabus Items for Exam Question Analysis .......................... 575
  John Foley and James Allan

Collaborative Filtering

Implicit Look-Alike Modelling in Display Ads – Transfer Collaborative Filtering to CTR Estimation ................................................................. 589
  Weinan Zhang, Lingxi Chen, and Jun Wang

Efficient Pseudo-Relevance Feedback Methods for Collaborative Filtering Recommendation ................................................................. 602
  Daniel Valcarce, Javier Parapar, and Álvaro Barreiro

Language Models for Collaborative Filtering Neighbourhoods ...................... 614
  Daniel Valcarce, Javier Parapar, and Álvaro Barreiro

Adaptive Collaborative Filtering with Extended Kalman Filters and Multi-armed Bandits ................................................................. 626
  Jean-Michel Renders

Short Papers

A Business Zone Recommender System Based on Facebook and Urban Planning Data. ................................................................. 641
  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 .............................. 648
  Walid Magdy, Tamer Elsayed, and Maram Hasanain

Finding Relevant Relations in Relevant Documents ..................................... 654
  Michael Schuhmacher, Benjamin Roth, Simone Paolo Ponzetto, and Laura Dietz

Probabilistic Multileave Gradient Descent ................................................. 661
  Harrie Oosterhuis, Anne Schuth, and Maarten de Rijke

Real-World Expertise Retrieval: The Information Seeking Behaviour of Recruitment Professionals .................................................. 669
  Tony Russell-Rose and Jon Chamberlain

Compressing and Decoding Term Statistics Time Series ........................ 675
  Jinfeng Rao, Xing Niu, and Jimmy Lin
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback or Research: Separating Pre-purchase from Post-purchase</td>
<td>682</td>
</tr>
<tr>
<td>Consumer Reviews</td>
<td></td>
</tr>
<tr>
<td>Mehedi Hasan, Alexander Kotov, Aravind Mohan, Shiyong Lu,</td>
<td></td>
</tr>
<tr>
<td>and Paul M. Stieg</td>
<td></td>
</tr>
<tr>
<td>Inferring the Socioeconomic Status of Social Media Users Based on</td>
<td>689</td>
</tr>
<tr>
<td>Behaviour and Language</td>
<td></td>
</tr>
<tr>
<td>Vasileios Lampos, Nikolaos Aletras, Jens K. Geyti, Bin Zou,</td>
<td></td>
</tr>
<tr>
<td>and Ingemar J. Cox</td>
<td></td>
</tr>
<tr>
<td>Two Scrolls or One Click: A Cost Model for Browsing Search Results</td>
<td>696</td>
</tr>
<tr>
<td>Leif Azzopardi and Guido Zuccon</td>
<td></td>
</tr>
<tr>
<td>Determining the Optimal Session Interval for Transaction Log Analysis</td>
<td>703</td>
</tr>
<tr>
<td>of an Online Library Catalogue</td>
<td></td>
</tr>
<tr>
<td>Simon Wakeling and Paul Clough</td>
<td></td>
</tr>
<tr>
<td>A Comparison of Deep Learning Based Query Expansion with Pseudo-Relevance Feedback and Mutual Information</td>
<td>709</td>
</tr>
<tr>
<td>Mohammad ALMasri, Catherine Berrut, and Jean-Pierre Chevallet</td>
<td></td>
</tr>
<tr>
<td>A Full-Text Learning to Rank Dataset for Medical Information Retrieval</td>
<td>716</td>
</tr>
<tr>
<td>Vera Boteva, Demian Gholipour, Artem Sokolov, and Stefan Riezler</td>
<td></td>
</tr>
<tr>
<td>Multi-label, Multi-class Classification Using Polylingual Embeddings</td>
<td>723</td>
</tr>
<tr>
<td>Georgios Balikas and Massih-Reza Amini</td>
<td></td>
</tr>
<tr>
<td>Learning Word Embeddings from Wikipedia for Content-Based Recommender Systems</td>
<td>729</td>
</tr>
<tr>
<td>Cataldo Musto, Giovanni Semeraro, Marco de Gemmis,</td>
<td></td>
</tr>
<tr>
<td>and Pasquale Lops</td>
<td></td>
</tr>
<tr>
<td>Tracking Interactions Across Business News, Social Media, and Stock</td>
<td>735</td>
</tr>
<tr>
<td>Fluctuations</td>
<td></td>
</tr>
<tr>
<td>Ossi Karkulahti, Lidia Pivovarova, Mian Du, Jussi Kangasharju,</td>
<td></td>
</tr>
<tr>
<td>and Roman Yangarber</td>
<td></td>
</tr>
<tr>
<td>Subtopic Mining Based on Three-Level Hierarchical Search Intentions</td>
<td>741</td>
</tr>
<tr>
<td>Se-Jong Kim, Jaehun Shin, and Jong-Hyeok Lee</td>
<td></td>
</tr>
<tr>
<td>Cold Start Cumulative Citation Recommendation for Knowledge Base</td>
<td>748</td>
</tr>
<tr>
<td>Acceleration</td>
<td></td>
</tr>
<tr>
<td>Jingang Wang, Jingtian Jiang, Lejian Liao, Dandan Song,</td>
<td></td>
</tr>
<tr>
<td>Zhiwei Zhang, and Chin-Yew Lin</td>
<td></td>
</tr>
<tr>
<td>Cross Domain User Engagement Evaluation</td>
<td>754</td>
</tr>
<tr>
<td>Ali Montazeralghaem, Hamed Zamani, and Azadeh Shakery</td>
<td></td>
</tr>
</tbody>
</table>
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

LExL: 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
TimeMachine: Entity-Centric Search and Visualization of News Archives . . . 845
  Pedro Saleiro, Jorge Teixeira, Carlos Soares, and Eugénio Oliveira

OPMES: A Similarity Search Engine for Mathematical Content . . . . . . . 849
  Wei Zhong and Hui Fang

SHAMUS: UFAL Search and Hyperlinking Multimedia System . . . . . . . 853
  Petra Galuščáková, Shadi Saleh, and Pavel Pecina

Industry Day

Industry Day Overview ................................................................. 859
  Omar Alonso and Pavel Serdyukov

Workshops

Bibliometric-Enhanced Information Retrieval: 3rd International BIR Workshop ................................................................. 865
  Philipp Mayr, Ingo Frommholz, and Guillaume Cabanac

MultiLingMine 2016: Modeling, Learning and Mining for Cross/Multilinguality ................................................................. 869
  Dino Ienco, Mathieu Roche, Salvatore Romeo, Paolo Rosso,
  and Andrea Tagarelli

Proactive Information Retrieval: Anticipating Users’ Information Need . . . 874
  Sumit Bhatia, Debapriyo Majumdar, and Nitish Aggarwal

First International Workshop on Recent Trends in News Information Retrieval (NewsIR’16) ................................................................. 878
  Miguel Martinez-Alvarez, Udo Kruschwitz, Gabriella Kazai,
  Frank Hopfgartner, David Corney, Ricardo Campos,
  and Dyaa Albakour

Tutorials

Collaborative Information Retrieval: Concepts, Models and Evaluation . . . 885
  Lynda Tamine and Laure Soulier

Group Recommender Systems: State of the Art, Emerging Aspects
and Techniques, and Research Challenges ........................................ 889
  Ludovico Boratto

Living Labs for Online Evaluation: From Theory to Practice .................. 893
  Anne Schuth and Krisztian Balog
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