

Nicola Ferro · Fabio Crestani  
Marie-Francine Moens · Josiane Mothe  
Fabrizio Silvestri · Giorgio Maria Di Nunzio  
Claudia Hauff · Gianmaria Silvello (Eds.)

# Advances in Information Retrieval

38th European Conference on IR Research, ECIR 2016  
Padua, Italy, March 20–23, 2016  
Proceedings

# Contents

## Social Context and News

SoRTESum: A Social Context Framework for Single-Document Summarization . . . . .	3
<i>Minh-Tien Nguyen and Minh-Le Nguyen</i>	
A Graph-Based Approach to Topic Clustering for Online Comments to News. . . . .	15
<i>Ahmet Aker, Emina Kurtic, A.R. Balamurali, Monica Paramita, Emma Barker, Mark Hepple, and Rob Gaizauskas</i>	
Leveraging Semantic Annotations to Link Wikipedia and News Archives . . .	30
<i>Arunav Mishra and Klaus Berberich</i>	

## Machine Learning

Deep Learning over Multi-field Categorical Data – A Case Study on User Response Prediction . . . . .	45
<i>Weinan Zhang, Tianming Du, and Jun Wang</i>	
Supervised Local Contexts Aggregation for Effective Session Search. . . . .	58
<i>Zhiwei Zhang, Jingang Wang, Tao Wu, Pengjie Ren, Zhumin Chen, and Luo Si</i>	
An Empirical Study of Skip-Gram Features and Regularization for Learning on Sentiment Analysis . . . . .	72
<i>Cheng Li, Bingyu Wang, Virgil Pavlu, and Javed A. Aslam</i>	
Multi-task Representation Learning for Demographic Prediction . . . . .	88
<i>Pengfei Wang, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng</i>	
Large-Scale Kernel-Based Language Learning Through the Ensemble Nystrom Methods . . . . .	100
<i>Danilo Croce and Roberto Basili</i>	

## Question Answering

Beyond Factoid QA: Effective Methods for Non-factoid Answer Sentence Retrieval . . . . .	115
<i>Liu Yang, Qingyao Ai, Damiano Spina, Ruey-Cheng Chen, Liang Pang, W. Bruce Croft, Jiafeng Guo, and Falk Scholer</i>	

Supporting Human Answers for Advice-Seeking Questions in CQA Sites . . .	129
<i>Liora Braunstain, Oren Kurland, David Carmel, Idan Szpektor, and Anna Shtok</i>	
<b>Ranking</b>	
Does Selective Search Benefit from WAND Optimization? . . . . .	145
<i>Yubin Kim, Jamie Callan, J. Shane Culpepper, and Alistair Moffat</i>	
Efficient AUC Optimization for Information Ranking Applications . . . . .	159
<i>Sean J. Welleck</i>	
Modeling User Interests for Zero-Query Ranking . . . . .	171
<i>Liu Yang, Qi Guo, Yang Song, Sha Meng, Milad Shokouhi, Kieran McDonald, and W. Bruce Croft</i>	
<b>Evaluation Methodology</b>	
Adaptive Effort for Search Evaluation Metrics . . . . .	187
<i>Jiepu Jiang and James Allan</i>	
Evaluating Memory Efficiency and Robustness of Word Embeddings . . . . .	200
<i>Johannes Jurgovsky, Michael Granitzer, and Christin Seifert</i>	
Characterizing Relevance on Mobile and Desktop . . . . .	212
<i>Manisha Verma and Emine Yilmaz</i>	
<b>Probabilistic Modelling</b>	
Probabilistic Local Expert Retrieval. . . . .	227
<i>Wen Li, Carsten Eickhoff, and Arjen P. de Vries</i>	
Probabilistic Topic Modelling with Semantic Graph . . . . .	240
<i>Long Chen, Joemon M. Jose, Haitao Yu, Fajie Yuan, and Huaizhi Zhang</i>	
Estimating Probability Density of Content Types for Promoting Medical Records Search. . . . .	252
<i>Yun He, Qinmin Hu, Yang Song, and Liang He</i>	
<b>Evaluation Issues</b>	
The Curious Incidence of Bias Corrections in the Pool . . . . .	267
<i>Aldo Lipani, Mihai Lupu, and Allan Hanbury</i>	
Understandability Biased Evaluation for Information Retrieval . . . . .	280
<i>Guido Zuccon</i>	

The Relationship Between User Perception and User Behaviour in Interactive Information Retrieval Evaluation . . . . .	293
<i>Mengdie Zhuang, Elaine G. Toms, and Gianluca Demartini</i>	

## Multimedia

Using Query Performance Predictors to Improve Spoken Queries . . . . .	309
<i>Jaime Arguello, Sandeep Avula, and Fernando Diaz</i>	
Fusing Web and Audio Predictors to Localize the Origin of Music Pieces for Geospatial Retrieval . . . . .	322
<i>Markus Schedl and Fang Zhou</i>	
Key Estimation in Electronic Dance Music . . . . .	335
<i>Ángel Faraldo, Emilia Gómez, Sergi Jordà, and Perfecto Herrera</i>	

## Summarization

Evaluating Text Summarization Systems with a Fair Baseline from Multiple Reference Summaries . . . . .	351
<i>Fahmida Hamid, David Haraburda, and Paul Tarau</i>	
Multi-document Summarization Based on Atomic Semantic Events and Their Temporal Relationships . . . . .	366
<i>Yllias Chali and Mohsin Uddin</i>	
Tweet Stream Summarization for Online Reputation Management . . . . .	378
<i>Jorge Carrillo-de-Albornoz, Enrique Amigó, Laura Plaza, and Julio Gonzalo</i>	

## Reproducibility

Who Wrote the Web? Revisiting Influential Author Identification Research Applicable to Information Retrieval . . . . .	393
<i>Martin Potthast, Sarah Braun, Tolga Buz, Fabian Duffhauss, Florian Friedrich, Jörg Marvin Gülzow, Jakob Köhler, Winfried Löttsch, Fabian Müller, Maïke Elisa Müller, Robert Paßmann, Bernhard Reinke, Lucas Rettenmeier, Thomas Rometsch, Timo Sommer, Michael Träger, Sebastian Wilhelm, Benno Stein, Efstathios Stamatatos, and Matthias Hagen</i>	
Toward Reproducible Baselines: The Open-Source IR Reproducibility Challenge . . . . .	408
<i>Jimmy Lin, Matt Crane, Andrew Trotman, Jamie Callan, Ishan Chattopadhyaya, John Foley, Grant Ingersoll, Craig Macdonald, and Sebastiano Vigna</i>	

Experiments in Newswire Summarisation . . . . .	421
<i>Stuart Mackie, Richard McCreadie, Craig Macdonald, and Iadh Ounis</i>	
On the Reproducibility of the TAGME Entity Linking System . . . . .	436
<i>Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg</i>	
<b>Twitter</b>	
Correlation Analysis of Reader’s Demographics and Tweet Credibility Perception . . . . .	453
<i>Shafiza Mohd Shariff, Mark Sanderson, and Xiuzhen Zhang</i>	
Topic-Specific Stylistic Variations for Opinion Retrieval on Twitter . . . . .	466
<i>Anastasia Giachanou, Morgan Harvey, and Fabio Crestani</i>	
Inferring Implicit Topical Interests on Twitter. . . . .	479
<i>Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri, and Mohsen Kahani</i>	
Topics in Tweets: A User Study of Topic Coherence Metrics for Twitter Data . . . . .	492
<i>Anjie Fang, Craig Macdonald, Iadh Ounis, and Philip Habel</i>	
<b>Retrieval Models</b>	
Supporting Scholarly Search with Keyqueries. . . . .	507
<i>Mathias Hagen, Anna Beyer, Tim Gollub, Kristof Komlossy, and Benno Stein</i>	
Pseudo-Query Reformulation . . . . .	521
<i>Fernando Diaz</i>	
VODUM: A Topic Model Unifying Viewpoint, Topic and Opinion Discovery. . . . .	533
<i>Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat</i>	
<b>Applications</b>	
Harvesting Training Images for Fine-Grained Object Categories Using Visual Descriptions . . . . .	549
<i>Josiah Wang, Katja Markert, and Mark Everingham</i>	
Do Your Social Profiles Reveal What Languages You Speak? Language Inference from Social Media Profiles. . . . .	561
<i>Yu Xu, M. Rami Ghorab, Zhongqing Wang, Dong Zhou, and Séamus Lawless</i>	

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*

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

An Empirical Comparison of Term Association and Knowledge Graphs for Query Expansion . . . . .	761
<i>Saeid Balaneshinkordan and Alexander Kotov</i>	
Deep Learning to Predict Patient Future Diseases from the Electronic Health Records . . . . .	768
<i>Riccardo Miotto, Li Li, and Joel T. Dudley</i>	
Improving Document Ranking for Long Queries with Nested Query Segmentation . . . . .	775
<i>Rishiraj Saha Roy, Anusha Suresh, Niloy Ganguly, and Monojit Choudhury</i>	
Sketching Techniques for Very Large Matrix Factorization. . . . .	782
<i>Raghavendran Balu, Teddy Furon, and Laurent Amsaleg</i>	
Diversifying Search Results Using Time: An Information Retrieval Method for Historians . . . . .	789
<i>Dhruv Gupta and Klaus Berberich</i>	
On Cross-Script Information Retrieval . . . . .	796
<i>Nada Naji and James Allan</i>	
LExL: A Learning Approach for Local Expert Discovery on Twitter . . . . .	803
<i>Wei Niu, Zhijiao Liu, and James Caverlee</i>	
Clickbait Detection . . . . .	810
<i>Martin Potthast, Sebastian Köpse, Benno Stein, and Matthias Hagen</i>	
Informativeness for Adhoc IR Evaluation: A Measure that Prevents Assessing Individual Documents . . . . .	818
<i>Romain Deveaud, Véronique Moriceau, Josiane Mothe, and Eric SanJuan</i>	
What Multimedia Sentiment Analysis Says About City Liveability . . . . .	824
<i>Joost Boonzajer Flaes, Stevan Rudinac, and Marcel Worring</i>	
<b>Demos</b>	
Scenemash: Multimodal Route Summarization for City Exploration. . . . .	833
<i>Jorrit van den Berg, Stevan Rudinac, and Marcel Worring</i>	
Exactus Like: Plagiarism Detection in Scientific Texts. . . . .	837
<i>Ilya Sochenkov, Denis Zubarev, Ilya Tikhomirov, Ivan Smirnov, Artem Shelmanov, Roman Suvorov, and Gennady Osipov</i>	
Jitter Search: A News-Based Real-Time Twitter Search Interface . . . . .	841
<i>Flávio Martins, João Magalhães, and Jamie Callan</i>	



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

Real-Time Bidding Based Display Advertising: Mechanisms and Algorithms . . . . .	897
<i>Jun Wang, Shuai Yuan, and Weinan Zhang</i>	
<b>Author Index</b> . . . . .	903