## Contents

Entity Linking to One Thousand Knowledge Bases ............................................ 1  
*Ning Gao and Silviu Cucerzan*

Where Do All These Search Terms Come From? – Two Experiments in Domain-Specific Search ................................................................. 15  
*Daniel Hienert and Maria Lusky*

Using Query Performance Predictors to Reduce Spoken Queries ..................... 27  
*Jaime Arguello, Sandeep Avula, and Fernando Diaz*

Entity Linking in Queries: Efficiency vs. Effectiveness ................................. 40  
*Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg*

Cross-Lingual Sentiment Relation Capturing for Cross-Lingual Sentiment Analysis ................................................................. 54  
*Qiang Chen, Wenjie Li, Yu Lei, Xule Liu, Chuwei Luo, and Yanxiang He*

Hierarchical Re-estimation of Topic Models for Measuring Topical Diversity ................................................................. 68  
*Hosein Azarbonyad, Mostafa Dehghani, Tom Kenter, Maarten Marx, Jaap Kamps, and Maarten de Rijke*

Collective Entity Linking in Tweets Over Space and Time .............................. 82  
*Wen-Haw Chong, Ee-Peng Lim, and William Cohen*

Simple Personalized Search Based on Long-Term Behavioral Signals ............. 95  
*Anna Sepliarskaia, Filip Radlinski, and Maarten de Rijke*

The Effects of Search Task Determinability on Search Behavior .................... 108  
*Rob Capra, Jaime Arguello, and Yinglong Zhang*

Inferring User Interests for Passive Users on Twitter by Leveraging Followee Biographies ................................................................. 122  
*Guangyuan Piao and John G. Breslin*

Fusion of Bag-of-Words Models for Image Classification in the Medical Domain ................................................................. 134  
*Leonidas Valavanis, Spyridon Stathopoulos, and Theodore Kalamboukis*

Do Topic Shift and Query Reformulation Patterns Correlate in Academic Search? ................................................................. 146  
*Xinyi Li and Maarten de Rijke*
Learning to Re-rank Medical Images Using a Bayesian Network-Based Thesaurus

Hajer Ayadi, Mouna Torjmen Khemakhem, Jimmy Xiangji Huang, Mariam Daoud, and Maher Ben Jemaa

A Part-of-Speech Enhanced Neural Conversation Model

Chuwei Luo, Wenjie Li, Qiang Chen, and Yanxiang He

“Are Machines Better Than Humans in Image Tagging?” - A User Study

Ralph Ewerth, Matthias Springstein, Lo An Phan-Vogtmann, and Juliane Schütze

Utterance Retrieval Based on Recurrent Surface Text Patterns

Guillaume Dubuisson Duplessis, Franck Charras, Vincent Letard, Anne-Laure Ligozat, and Sophie Rosset

How Do Order and Proximity Impact the Readability of Event Summaries?

Arunav Mishra and Klaus Berberich

Sentiment Propagation for Predicting Reputation Polarity

Anastasia Giachanou, Julio Gonzalo, Ida Mele, and Fabio Crestani

Transitivity, Time Consumption, and Quality of Preference Judgments in Crowdsourcing

Kai Hui and Klaus Berberich

Exploring Time-Sensitive Variational Bayesian Inference LDA for Social Media Data

Anjie Fang, Craig Macdonald, Iadh Ounis, Philip Habel, and Xiao Yang


David Brazier and Morgan Harvey

A Task Completion Engine to Enhance Search Session Support for Air Traffic Work Tasks

Yashar Moshfeghi, Raoul Rothfeld, Leif Azzopardi, and Peter Triantafillou

Personalized Keyword Boosting for Venue Suggestion Based on Multiple LBSNs

Mohammad Aliannejadi, Dimitrios Rafailidis, and Fabio Crestani

Feature-Oriented Analysis of User Profile Completion Problem

Morteza Haghir Chehreghani
Generating Descriptions of Entity Relationships .............................. 317
Nikos Voskarides, Edgar Meij, and Maarten de Rijke

Language Influences on Tweeter Geolocation .............................. 331
Ahmed Mourad, Falk Scholer, and Mark Sanderson

Human-Based Query Difficulty Prediction .............................. 343
Adrian-Gabriel Chifu, Sébastien Déjean, Stefano Mizzaro,
and Josiane Mothe

Fixed-Cost Pooling Strategies Based on IR Evaluation Measures .............. 357
Aldo Lipani, Joao Palotti, Mihai Lupu, Florina Piroi, Guido Zuccon,
and Allan Hanbury

A Formal and Empirical Study of Unsupervised Signal Combination
for Textual Similarity Tasks ............................................ 369
Enrique Amigó, Fernando Giner, Julio Gonzalo, and Felisa Verdejo

Paper2vec: Combining Graph and Text Information for Scientific
Paper Representation .............................................. 383
Soumyajit Ganguly and Vikram Pudi

Exploration of a Threshold for Similarity Based on Uncertainty
in Word Embedding ............................................ 396
Navid Rekabsaz, Mihai Lupu, and Allan Hanbury

Ancient Roman Coin Retrieval: A Systematic Examination of the Effects
of Coin Grade ............................................. 410
Callum Fare and Ognjen Arandjelović

A Systematic Analysis of Sentence Update Detection
for Temporal Summarization ............................................ 424
Cristina Gârbacea and Evangelos Kanoulas

A Multiple-Instance Learning Approach to Sentence Selection
for Question Ranking ............................................. 437
Salvatore Romeo, Giovanni Da San Martino, Alberto Barrón-Cedeño,
and Alessandro Moschitti

Enhancing Sensitivity Classification with Semantic Features
Using Word Embeddings ............................................. 450
Graham McDonald, Craig Macdonald, and Iadh Ounis

Predicting Users’ Future Interests on Twitter .............................. 464
Fattane Zarrinkalam, Hossein Fani, Ebrahim Bagheri,
and Mohsen Kahani
A New Scheme for Scoring Phrases in Unsupervised Keyphrase Extraction . . . 477
   Corina Florescu and Cornelia Caragea

JustEvents: A Crowdsourced Corpus for Event Validation with Strict
Temporal Constraints .......................................................... 484
   Andrea Ceroni, Ujwal Gadiraju, and Marco Fisichella

Dimension Projection Among Languages Based on Pseudo-Relevant
Documents for Query Translation ........................................ 493
   Javid Dadashkarimi, Mahsa S. Shahshahani, Amirhossein Tebbifakhr,
   Heshaam Faili, and Azadeh Shakery

Labeling Topics with Images Using a Neural Network ............ 500
   Nikolaos Aletras and Arpit Mittal

Leveraging Site Search Logs to Identify Missing Content
on Enterprise Webpages ..................................................... 506
   Harsh Jhamtani, Rishiraj Saha Roy, Niyati Chhaya, and Eric Nyberg

LTRo: Learning to Route Queries in Clustered P2P IR ............. 513
   Rami S. Alkhawaldeh, Deepak P., Joemon M. Jose, and Fajie Yuan

Faster K-Means Cluster Estimation ..................................... 520
   Siddhesh Khandelwal and Amit Awekar

Predicting Emotional Reaction in Social Networks ............... 527
   Jérémie Clos, Anil Bandhakavi, Nirmalie Wiratunga,
   and Guillaume Cabanac

Irony Detection with Attentive Recurrent Neural Networks .......... 534
   Yu-Hsiang Huang, Hen-Hsen Huang, and Hsin-Hsi Chen

We Used Neural Networks to Detect Clickbaits: You Won’t Believe
What Happened Next! ...................................................... 541
   Ankesh Anand, Tanmoy Chakraborty, and Noseong Park

Learning to Classify Inappropriate Query-Completions .......... 548
   Parth Gupta and Jose Santos

Counteracting Novelty Decay in First Story Detection ............ 555
   Yumeng Qin, Dominik Wurzer, Victor Lavrenko, and Cunchen Tang

Predicting Genre Preferences from Cultural and Socio-Economic Factors
for Music Retrieval .......................................................... 561
   Marcin Skowron, Florian Lemmerich, Bruce Ferwerda,
   and Markus Schedl
Batch Incremental Shared Nearest Neighbor Density Based Clustering
Algorithm for Dynamic Datasets ....................................................... 568
  Panthadeep Bhattacharjee and Amit Awekar

Improving Tweet Representations Using Temporal and User Context .... 575
  J. Ganesh, Manish Gupta, and Vasudeva Varma

Personalized Parsimonious Language Models for User Modeling in Social
Bookmaking Systems ................................................................. 582
  Nawal Ould Amer, Philippe Mulhem, and Mathias Géry

A Novel Word Embedding Based Stemming Approach for Microblog
Retrieval During Disasters .......................................................... 589
  Moumita Basu, Anurag Roy, Kripabandhu Ghosh,
  Somprakash Bandyopadhyay, and Saptarshi Ghosh

Search Personalization with Embeddings ....................................... 598
  Thanh Vu, Dat Quoc Nguyen, Mark Johnson, Dawei Song,
  and Alistair Willis

Do Easy Topics Predict Effectiveness Better Than Difficult Topics? ...... 605
  Kevin Roitero, Eddy Maddalena, and Stefano Mizzaro

Named Entity Classification Using Search Engine’s Query Suggestions ... 612
  Jayendra Barua and Dhaval Patel

“When Was This Picture Taken?” – Image Date Estimation in the Wild .... 619
  Eric Müller, Matthias Springstein, and Ralph Ewerth

Low-Cost Preference Judgment via Ties ......................................... 626
  Kai Hui and Klaus Berberich

Using Section Headings to Compute Cross-Lingual Similarity
of Wikipedia Articles ............................................................... 633
  Monica Lestari Paramita, Paul Clough, and Robert Gaizauskas

Learning to Rank for Consumer Health Search: A Semantic Approach ... 640
  Luca Soldaini and Nazli Goharian

Matrix Factorisation with Word Embeddings for Rating Prediction on
Location-Based Social Networks .................................................. 647
  Jarana Manotumruksa, Craig Macdonald, and Iadh Ounis

Word Similarity Based Model for Tweet Stream Prospective Notification. 655
  Abdelhamid Chellal, Mohand Boughanem, and Bernard Dousset
Negative Feedback in the Language Modeling Framework for Text Recommendation

Hossein Rahmatizadeh Zaghehi, Mozdeh Ariannezhad, and Azadeh Shakery

662

Plagiarism Detection in Texts Obsfuscated with Homoglyphs

Faisal Alvi, Mark Stevenson, and Paul Clough

669

Iterative Estimation of Document Relevance Score for Pseudo-Relevance Feedback

Mozdeh Ariannezhad, Ali Montazeralghaem, Hamed Zamani, and Azadeh Shakery

676

Design Patterns for Fusion-Based Object Retrieval

Shuo Zhang and Krisztian Balog

684

On the Long-Tail Entities in News

José Esquivel, Dyaa Albakour, Miguel Martinez, David Corney, and Samir Moussa

691

Search Costs vs. User Satisfaction on Mobile

Manisha Verma and Emine Yilmaz

698

On the Efficiency of Selective Search

Fatih Hafizoglu, Emre Can Kucukoglu, and Ismail Sengor Altingovde

705

Does Online Evaluation Correspond to Offline Evaluation in Query Auto Completion?

Alexandros Bampoulidis, João Palotti, Mihai Lupu, Jon Brassey, and Allan Hanbury

713

A Neural Attention Model for Categorizing Patient Safety Events

Arman Cohan, Allan Fong, Nazli Goharian, and Raj Ratwani

720

Promoting Understandability in Consumer Health Information Search

Hua Yang and Teresa Goncalves

727

A Social Framework for Set Recommendation in Group Recommender Systems

Lorena Recalde

735

Pyndri: A Python Interface to the Indri Search Engine

Christophe Van Gysel, Evangelos Kanoulas, and Maarten de Rijke

744

“Hey, vitrivr!” – A Multimodal UI for Video Retrieval

Prateek Goel, Ivan Giangreco, Luca Rossetto, Claudiu Tănase, and Hetko Schuldt

749
FairScholar: Balancing Relevance and Diversity for Scientific Paper Recommendation .................................................. 753
Ankesh Anand, Tanmoy Chakraborty, and Amitava Das

The SENSEI Overview of Newspaper Readers’ Comments ............ 758
Adam Funk, Ahmet Aker, Emma Barker, Monica Lestari Paramita, Mark Hepple, and Robert Gaizauskas

Temporal Semantic Analysis of Conference Proceedings ................. 762
Fedelucio Narducci, Pierpaolo Basile, Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro

QweetFinder: Real-Time Finding and Filtering of Question Tweets .... 766
Ameer Albahem, Maram Hasanain, Marwan Torki, and Tamer Elsayed

Integration of the Scientific Recommender System Mr. DLlib into the Reference Manager JabRef ............................. 770
Stefan Feyer, Sophie Siebert, Bela Gipp, Akiko Aizawa, and Joeran Beel

Abstracts of Doctoral Consortium Papers and Workshops

A Social Framework for Set Recommendation in Group Recommender Systems ..................................................... 777
Lorena Recalde

User Behavior Analysis and User Modeling for Complex Search ....... 778
Jiaxin Mao

First International Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) ......................... 779
Saptarshi Ghosh, Kripabandhu Ghosh, Tanmoy Chakraborty, Debasis Ganguly, Gareth Jones, and Marie-Francine Moens

Bibliometric-Enhanced Information Retrieval: 5th International BIR Workshop ...................................................... 784
Philipp Mayr, Ingo Frommholz, and Guillaume Cabanac

OnST’17: The 2nd International Workshop on Online Safety Trust Fraud Prevention ..................................................... 790
Marco Fisichella

Workshop on Social Media for Personalization and Search (SoMePeAS) .... 795
Ludovico Boratto, Andreas Kaltenbrunner, and Giovanni Stilo

Author Index ........................................................................ 797
Advances in Information Retrieval
39th European Conference on IR Research, ECIR 2017,
Aberdeen, UK, April 8-13, 2017, Proceedings
Joemon M., J.; Hauff, C.; Altingövde, I.S.; Song, D.;
Albakour, D.; Watt, S.; Tait, J. (Eds.)
2017, XXXI, 799 p. 146 illus., Softcover
ISBN: 978-3-319-56607-8