# Contents

## Efficiency

On Structures of Inverted Index for Query Processing Efficiency .......................... 3  
*Xingshen Song, Xueping Zhang, Yuexiang Yang, Jicheng Quan, and Kun Jiang*

Access Time Tradeoffs in Archive Compression ...................................................... 15  
*Matthias Petri, Alistair Moffat, P.C. Nagesh, and Anthony Wirth*

Large Scale Sentiment Analysis with Locality Sensitive BitHash ......................... 29  
*Wenhao Zhang, Jianqiu Ji, Jun Zhu, Hua Xu, and Bo Zhang*

## Graphs, Knowledge Bases and Taxonomies

Knowledge-Based Query Expansion in Real-Time Microblog Search ....................... 43  
*Chao Lv, Runwei Qiang, Feifan Fan, and Jianwu Yang*

Enrichment of Academic Search Engine Results Pages  
by Citation-Based Graphs ....................................................................................... 56  
*Shuhei Shogen, Toshiyuki Shimizu, and Masatoshi Yoshikawa*

Quality-Aware Review Selection Based on Product Feature Taxonomy .................. 68  
*Nan Tian, Yue Xu, Yuefeng Li, and Gabriella Pasi*

## Recommendations

An Author Subject Topic Model for Expert Recommendation .................................. 83  
*Haikun Mou, Qian Geng, Jian Jin, and Chong Chen*

A Keyword Recommendation Method Using CorKeD Words  
and Its Application to Earth Science Data ............................................................. 96  
*Youichi Ishida, Toshiyuki Shimizu, and Masatoshi Yoshikawa*

Incorporating Distinct Opinions in Content Recommender System ....................... 109  
*Grace E. Lee, Keejun Han, and Mun Y. Yi*

## Twitter and Social Media

Detecting Automatically-Generated Arabic Tweets ................................................. 123  
*Hind Almerekhi and Tamer Elsayed*
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving Tweet Timeline Generation by Predicting Optimal Retrieval Depth</td>
<td>135</td>
</tr>
<tr>
<td><em>Maram Hasanain, Tamer Elsayed, and Walid Magdy</em></td>
<td></td>
</tr>
<tr>
<td>Abstract Venue Concept Detection from Location-Based Social Networks</td>
<td>147</td>
</tr>
<tr>
<td><em>Yi Liao, Shoaib Jameel, Wai Lam, and Xing Xie</em></td>
<td></td>
</tr>
<tr>
<td><strong>Web Search</strong></td>
<td></td>
</tr>
<tr>
<td>Snapboard: A Shared Space of Visual Snippets - A Study in Individual and Asynchronous Collaborative Web Search</td>
<td>161</td>
</tr>
<tr>
<td><em>Teerapong Leelanupab, Hannarin Kruajirayu, and Nont Kanungsukkasem</em></td>
<td></td>
</tr>
<tr>
<td>Improving Ranking and Robustness of Search Systems by Exploiting the Popularity of Documents</td>
<td>174</td>
</tr>
<tr>
<td><em>Ashraf Bah and Ben Carterette</em></td>
<td></td>
</tr>
<tr>
<td>Heading-Aware Snippet Generation for Web Search</td>
<td>188</td>
</tr>
<tr>
<td><em>Tomohiro Manabe and Keishi Tajima</em></td>
<td></td>
</tr>
<tr>
<td><strong>Text Processing, Understanding and Categorization</strong></td>
<td></td>
</tr>
<tr>
<td>Smoothing Temporal Difference for Text Categorization</td>
<td>203</td>
</tr>
<tr>
<td><em>Fumiyo Fukumoto and Yoshimi Suzuki</em></td>
<td></td>
</tr>
<tr>
<td>Charset Encoding Detection of HTML Documents: A Practical Experience</td>
<td>215</td>
</tr>
<tr>
<td><em>Shabanali Faghani, Ali Hadian, and Behrouz Minaei-Bidgoli</em></td>
<td></td>
</tr>
<tr>
<td>Structure Matters: Adoption of Structured Classification Approach in the Context of Cognitive Presence Classification</td>
<td>227</td>
</tr>
<tr>
<td><em>Zak Waters, Vitomir Kovanović, Kirsty Kitto, and Dragan Gašević</em></td>
<td></td>
</tr>
<tr>
<td><strong>Topics and Models</strong></td>
<td></td>
</tr>
<tr>
<td>A Sequential Latent Topic-Based Readability Model for Domain-Specific Information Retrieval</td>
<td>241</td>
</tr>
<tr>
<td><em>Wenya Zhang, Dawei Song, Peng Zhang, Xiaozhao Zhao, and Yuexian Hou</em></td>
<td></td>
</tr>
<tr>
<td>Automatic Labelling of Topic Models Using Word Vectors and Letter Trigram Vectors</td>
<td>253</td>
</tr>
<tr>
<td><em>Wanqiu Kou, Fang Li, and Timothy Baldwin</em></td>
<td></td>
</tr>
<tr>
<td>A Study of Collection-Based Features for Adapting the Balance Parameter in Pseudo Relevance Feedback</td>
<td>265</td>
</tr>
<tr>
<td><em>Ye Meng, Peng Zhang, Dawei Song, and Yuexian Hou</em></td>
<td></td>
</tr>
</tbody>
</table>
Clustering

An MDL-Based Frequent Itemset Hierarchical Clustering Technique to Improve Query Search Results of an Individual Search Engine ............ 279
   Diyah Puspitaningrum, Fauzi, Boko Susilo, Jeri Apriansyah Pagua, Aan Erlansari, Desi Andreswari, Rusdi Efendi, and I.S.W.B. Prasetya

Improving Clustering Quality by Automatic Text Summarization ............. 292
   Mohsen Pourvali, Salvatore Orlando, and Mehrad Gharagozloo

Tweet Timeline Generation via Graph-Based Dynamic Greedy Clustering ... 304
   Feifan Fan, Runwei Qiang, Chao Lv, Wayne Xin Zhao, and Jianwu Yang

Evaluation

Topic Set Size Design with the Evaluation Measures for Short Text Conversation ................................................................. 319
   Tetsuya Sakai, Lifeng Shang, Zhengdong Lu, and Hang Li

Towards Nuanced System Evaluation Based on Implicit User Expectations ... 332
   Paul Thomas, Peter Bailey, Alistair Moffat, and Falk Scholer

Social Media and Recommendation

   Yangjie Yao and Aixin Sun

Company Name Disambiguation in Tweets: A Two-Step Filtering Approach ................................................................. 358
   M. Atif Qureshi, Arjumand Younus, Colm O’Riordan, and Gabriella Pasi

Utilizing Word Embeddings for Result Diversification in Tweet Search ........ 366
   Kezban Dilek Onal, Ismail Sengor Altingovde, and Pinar Karagoz

Short Papers

Displaying People with Old Addresses on a Map ......................... 381
   Gang Zhang and Harumi Murakami

A Query Expansion Approach Using Entity Distribution Based on Markov Random Fields ................................................................. 387
   Rui Li, Linxue Hao, Xiaozhao Zhao, Peng Zhang, Dawei Song, and Yuexian Hou
Analysis of Cyber Army’s Behaviours on Web Forum for Elect Campaign ... 394
   Man-Chun Ko and Hsin-Hsi Chen

Beyond tf-idf and Cosine Distance in Documents Dissimilarity Measure ... 400
   Sunil Aryal, Kai Ming Ting, Gholamreza Haffari, and Takashi Washio

Explorations of Cross-Disciplinary Term Similarity ...................... 407
   Hanif Sheikhadbolkarim and Laurianne Sitbon

A Signature Approach to Patent Classification .......................... 413
   Dilesha Seneviratne, Shlomo Geva, Guido Zuccon, Gabriela Ferraro,
   Timothy Chappell, and Magali Meireles

The Impact of Using Combinatorial Optimisation for Static Caching
of Posting Lists. ................................................................. 420
   Casper Petersen, Jakob Grue Simonsen, and Christina Lioma

Improving Difficult Queries by Leveraging Clusters in Term Graph ...... 426
   Rajul Anand and Alexander Kotov

Demonstrations

EEST: Entity-Driven Exploratory Search for Twitter ....................... 435
   Chao Lv, Runwei Qiang, Lili Yao, and Jianwu Yang

Oyster: A Tool for Fine-Grained Ontological Annotations in Free-Text ... 440
   Hamed Tayebikhorami, Alejandro Metke-Jimenez, Anthony Nguyen,
   and Guido Zuccon

TopSig: A Scalable System for Hashing and Retrieving
Document Signatures .......................................................... 447
   Timothy Chappell and Shlomo Geva

Author Index ........................................................................ 453
Information Retrieval Technology
Zuccon, G.; Geva, S.; Joho, H.; Scholer, F.; Sun, A.; Zhang, P. (Eds.)
2015, XVIII, 454 p. 84 illus. in color., Softcover
ISBN: 978-3-319-28939-7