## 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
Improving Tweet Timeline Generation by Predicting Optimal Retrieval Depth. ................................................................. 135
  Maram Hasanain, Tamer Elsayed, and Walid Magdy

Abstract Venue Concept Detection from Location-Based Social Networks ... 147
  Yi Liao, Shoaib Jameel, Wai Lam, and Xing Xie

Web Search

Snapboard: A Shared Space of Visual Snippets - A Study in Individual and Asynchronous Collaborative Web Search ................................. 161
  Teerapong Leelanupab, Hannarin Kruajirayu, and Nont Kanungsukkasem

Improving Ranking and Robustness of Search Systems by Exploiting the Popularity of Documents ..................................................... 174
  Ashraf Bah and Ben Carterette

Heading-Aware Snippet Generation for Web Search ......................... 188
  Tomohiro Manabe and Keishi Tajima

Text Processing, Understanding and Categorization

Smoothing Temporal Difference for Text Categorization ....................... 203
  Fumiyo Fukumoto and Yoshimi Suzuki

Charset Encoding Detection of HTML Documents: A Practical Experience . .... 215
  Shabanali Faghani, Ali Hadian, and Behrouz Minai-Bidgoli

Structure Matters: Adoption of Structured Classification Approach in the Context of Cognitive Presence Classification ......................... 227
  Zak Waters, Vitomir Kovanović, Kirsty Kitto, and Dragan Gašević

Topics and Models

A Sequential Latent Topic-Based Readability Model for Domain-Specific Information Retrieval ........................................................... 241
  Wenya Zhang, Dawei Song, Peng Zhang, Xiaozhao Zhao, and Yuexian Hou

Automatic Labelling of Topic Models Using Word Vectors and Letter Trigram Vectors ................................................................. 253
  Wanqiu Kou, Fang Li, and Timothy Baldwin

A Study of Collection-Based Features for Adapting the Balance Parameter in Pseudo Relevance Feedback ........................................ 265
  Ye Meng, Peng Zhang, Dawei Song, and Yuexian Hou
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td></td>
</tr>
<tr>
<td>An MDL-Based Frequent Itemset Hierarchical Clustering Technique</td>
<td>279</td>
</tr>
<tr>
<td>to Improve Query Search Results of an Individual Search Engine</td>
<td></td>
</tr>
<tr>
<td><em>Diyah Puspitaningrum, Fauzi, Boko Susilo, Jeri Apriansyah Pagua,</em></td>
<td></td>
</tr>
<tr>
<td><em>Aan Erlansari, Desi Andreswari, Rusdi Efendi, and I.S.W.B. Prasetya</em></td>
<td></td>
</tr>
<tr>
<td>Improving Clustering Quality by Automatic Text Summarization</td>
<td>292</td>
</tr>
<tr>
<td><em>Mohsen Pourvali, Salvatore Orlando, and Mehrad Gharagozloo</em></td>
<td></td>
</tr>
<tr>
<td>Tweet Timeline Generation via Graph-Based Dynamic Greedy Clustering</td>
<td>304</td>
</tr>
<tr>
<td><em>Feifan Fan, Runwei Qiang, Chao Lv, Wayne Xin Zhao, and Jianwu Yang</em></td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
</tr>
<tr>
<td>Topic Set Size Design with the Evaluation Measures for Short Text</td>
<td>319</td>
</tr>
<tr>
<td>Conversation</td>
<td></td>
</tr>
<tr>
<td><em>Tetsuya Sakai, Lifeng Shang, Zhengdong Lu, and Hang Li</em></td>
<td></td>
</tr>
<tr>
<td>Towards Nuanced System Evaluation Based on Implicit User Expectations</td>
<td>332</td>
</tr>
<tr>
<td><em>Paul Thomas, Peter Bailey, Alistair Moffat, and Falk Scholer</em></td>
<td></td>
</tr>
<tr>
<td>Social Media and Recommendation</td>
<td></td>
</tr>
<tr>
<td>A Study of Visual and Semantic Similarity for Social Image Search</td>
<td>347</td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
</tr>
<tr>
<td><em>Yangjie Yao and Aixin Sun</em></td>
<td></td>
</tr>
<tr>
<td>Company Name Disambiguation in Tweets: A Two-Step Filtering Approach</td>
<td>358</td>
</tr>
<tr>
<td><em>M. Atif Qureshi, Arjumand Younus, Colm O’Riordan,</em></td>
<td></td>
</tr>
<tr>
<td><em>and Gabriella Pasi</em></td>
<td></td>
</tr>
<tr>
<td>Utilizing Word Embeddings for Result Diversification in Tweet Search</td>
<td>366</td>
</tr>
<tr>
<td><em>Kezban Dilek Onal, Ismail Sengor Altingovde, and Pinar Karagoz</em></td>
<td></td>
</tr>
<tr>
<td>Short Papers</td>
<td></td>
</tr>
<tr>
<td>Displaying People with Old Addresses on a Map</td>
<td>381</td>
</tr>
<tr>
<td><em>Gang Zhang and Harumi Murakami</em></td>
<td></td>
</tr>
<tr>
<td>A Query Expansion Approach Using Entity Distribution Based on</td>
<td>387</td>
</tr>
<tr>
<td>Markov Random Fields</td>
<td></td>
</tr>
<tr>
<td><em>Rui Li, Linxue Hao, Xiaozhao Zhao, Peng Zhang, Dawei Song,</em></td>
<td></td>
</tr>
<tr>
<td><em>and Yuexian Hou</em></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Analysis of Cyber Army’s Behaviours on Web Forum for Elect Campaign</td>
<td>394</td>
</tr>
<tr>
<td>Man-Chun Ko and Hsin-Hsi Chen</td>
<td></td>
</tr>
<tr>
<td>Beyond tf-idf and Cosine Distance in Documents Dissimilarity Measure</td>
<td>400</td>
</tr>
<tr>
<td>Sunil Aryal, Kai Ming Ting, Gholamreza Haffari, and Takashi Washio</td>
<td></td>
</tr>
<tr>
<td>Explorations of Cross-Disciplinary Term Similarity</td>
<td>407</td>
</tr>
<tr>
<td>Hanif Sheikhadbolkarim and Laurianne Sitbon</td>
<td></td>
</tr>
<tr>
<td>A Signature Approach to Patent Classification</td>
<td>413</td>
</tr>
<tr>
<td>Dileshe Seneviratne, Shlomo Geva, Guido Zuccon, Gabriela Ferraro,</td>
<td></td>
</tr>
<tr>
<td>Timothy Chappell, and Magali Meireles</td>
<td></td>
</tr>
<tr>
<td>The Impact of Using Combinatorial Optimisation for Static Caching</td>
<td>420</td>
</tr>
<tr>
<td>Casper Petersen, Jakob Grue Simonsen, and Christina Lioma</td>
<td></td>
</tr>
<tr>
<td>Improving Difficult Queries by Leveraging Clusters in Term Graph</td>
<td>426</td>
</tr>
<tr>
<td>Rajul Anand and Alexander Kotov</td>
<td></td>
</tr>
<tr>
<td><strong>Demonstrations</strong></td>
<td></td>
</tr>
<tr>
<td>EEST: Entity-Driven Exploratory Search for Twitter</td>
<td>435</td>
</tr>
<tr>
<td>Chao Lv, Runwei Qiang, Lili Yao, and Jianwu Yang</td>
<td></td>
</tr>
<tr>
<td>Oyster: A Tool for Fine-Grained Ontological Annotations in Free-Text</td>
<td>440</td>
</tr>
<tr>
<td>Hamed Tayebikhorami, Alejandro Metke-Jimenez, Anthony Nguyen,</td>
<td></td>
</tr>
<tr>
<td>and Guido Zuccon</td>
<td></td>
</tr>
<tr>
<td>TopSig: A Scalable System for Hashing and Retrieving Document Signs</td>
<td>447</td>
</tr>
<tr>
<td>Timothy Chappell and Shlomo Geva</td>
<td></td>
</tr>
<tr>
<td><strong>Author Index</strong></td>
<td>453</td>
</tr>
</tbody>
</table>
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