

Fabio Crestani Paolo Ferragina
Mark Sanderson (Eds.)

String Processing and Information Retrieval

13th International Conference, SPIRE 2006
Glasgow, UK, October 11-13, 2006
Proceedings

 Springer

Table of Contents

Web Clustering and Text Categorization

MP-Boost: A Multiple-Pivot Boosting Algorithm and Its Application to Text Categorization	1
<i>Andrea Esuli, Tiziano Fagni, Fabrizio Sebastiani</i>	
TreeBoost.MH: A Boosting Algorithm for Multi-label Hierarchical Text Categorization	13
<i>Andrea Esuli, Tiziano Fagni, Fabrizio Sebastiani</i>	
Cluster Generation and Cluster Labelling for Web Snippets	25
<i>Filippo Geraci, Marco Pellegrini, Marco Maggini, Fabrizio Sebastiani</i>	
Principal Components for Automatic Term Hierarchy Building.....	37
<i>Georges Dupret, Benjamin Piwowarski</i>	

Strings

Computing the Minimum Approximate λ -Cover of a String.....	49
<i>Qing Guo, Hui Zhang, Costas S. Iliopoulos</i>	
Sparse Directed Acyclic Word Graphs	61
<i>Shunsuke Inenaga, Masayuki Takeda</i>	
On-Line Repetition Detection	74
<i>Jin-Ju Hong, Gen-Huey Chen</i>	

User Behavior

Analyzing User Behavior to Rank Desktop Items.....	86
<i>Paul-Alexandru Chirita, Wolfgang Nejdl</i>	
The Intention Behind Web Queries	98
<i>Ricardo Baeza-Yates, Liliana Calderón-Benavides, Cristina González-Caro</i>	

Web Search Algorithms

Compact Features for Detection of Near-Duplicates in Distributed Retrieval	110
<i>Yaniv Bernstein, Milad Shokouhi, Justin Zobel</i>	
Inverted Files Versus Suffix Arrays for Locating Patterns in Primary Memory	122
<i>Simon J. Puglisi, W.F. Smyth, Andrew Turpin</i>	
Efficient Lazy Algorithms for Minimal-Interval Semantics	134
<i>Paolo Boldi, Sebastiano Vigna</i>	
Output-Sensitive Autocompletion Search	150
<i>Holger Bast, Christian W. Mortensen, Ingmar Weber</i>	

Compression

A Compressed Self-index Using a Ziv-Lempel Dictionary	163
<i>Luís M.S. Russo, Arlindo L. Oliveira</i>	
Mapping Words into Codewords on PPM	181
<i>Joaquín Adiego, Pablo de la Fuente</i>	

Correction

Improving Usability Through Password-Corrective Hashing	193
<i>Andrew Mehler, Steven Skiena</i>	
Word-Based Correction for Retrieval of Arabic OCR Degraded Documents	205
<i>Walid Magdy, Kareem Darwish</i>	

Information Retrieval Applications

A Statistical Model of Query Log Generation	217
<i>Georges Dupret, Benjamin Piwowarski, Carlos Hurtado, Marcelo Mendoza</i>	
Using String Comparison in Context for Improved Relevance Feedback in Different Text Media	229
<i>Adenike M. Lam-Adesina, Gareth J.F. Jones</i>	

A Multiple Criteria Approach for Information Retrieval	242
<i>Mohamed Farah, Daniel Vanderpooten</i>	

English to Persian Transliteration	255
<i>Sarvnaz Karimi, Andrew Turpin, Falk Scholer</i>	

Bio Informatics

Efficient Algorithms for Pattern Matching with General Gaps and Character Classes	267
<i>Kimmo Fredriksson, Szymon Grabowski</i>	

Matrix Tightness: A Linear-Algebraic Framework for Sorting by Transpositions	279
<i>Tzvika Hartman, Elad Verbin</i>	

How to Compare Arc-Annotated Sequences: The Alignment Hierarchy	291
<i>Guillaume Blin, H�el�ene Touzet</i>	

Web Search Engines

Structured Index Organizations for High-Throughput Text Querying	304
<i>Vo Ngoc Anh, Alistair Moffat</i>	

Adaptive Query-Based Sampling of Distributed Collections	316
<i>Mark Baillie, Leif Azzopardi, Fabio Crestani</i>	

Short Papers

Dotted Suffix Trees A Structure for Approximate Text Indexing	329
<i>Lu�s Pedro Coelho, Artindo L. Oliveira</i>	

Phrase-Based Pattern Matching in Compressed Text	337
<i>J. Shane Culpepper, Alistair Moffat</i>	

Discovering Context-Topic Rules in Search Engine Logs	346
<i>Carlos A. Hurlado, Mark Levene</i>	

Incremental Aggregation of Latent Semantics Using a Graph-Based Energy Model	354
<i>Aditya Ramana Rachakonda, Srinath Srinivasa</i>	

XIV Table of Contents

A New Algorithm for Fast All-Against-All Substring Matching 360
Marina Barsky, Ulrike Stege, Alex Thomo, Chris Upton

Author Index 367