

Table of Contents

Efficient One Dimensional Real Scaled Matching	1
<i>Amihood Amir, Ayelet Butman, Moshe Lewenstein, Ely Porat, and Dekel Tsur</i>	
Linear Time Algorithm for the Longest Common Repeat Problem	10
<i>Inbok Lee, Costas S. Iliopoulos, and Kunsoo Park</i>	
Automaton-Based Sublinear Keyword Pattern Matching	18
<i>Loek Cleophas, Bruce W. Watson, and Gerard Zwaan</i>	
Techniques for Efficient Query Expansion	30
<i>Bodo Billerbeck and Justin Zobel</i>	
Inferring Query Performance Using Pre-retrieval Predictors	43
<i>Ben He and Iadh Ounis</i>	
A Scalable System for Identifying Co-derivative Documents	55
<i>Yaniv Bernstein and Justin Zobel</i>	
Searching for a Set of Correlated Patterns	68
<i>Shmuel T. Klein and Riva Shalom</i>	
Linear Nondeterministic Dawg String Matching Algorithm	70
<i>Longtao He and Binxing Fang</i>	
Permuted and Scaled String Matching	72
<i>Ayelet Butman, Revital Eres, and Gad M. Landau</i>	
Bit-Parallel Branch and Bound Algorithm for Transposition Invariant LCS	74
<i>Kjell Lemström, Gonzalo Navarro, and Yoan Pinzon</i>	
A New Feature Normalization Scheme Based on Eigenspace for Noisy Speech Recognition	76
<i>Yoonjae Lee and Hanseok Ko</i>	
Fast Detection of Common Sequence Structure Patterns in RNAs	79
<i>Rolf Backofen and Sven Siebert</i>	
An Efficient Algorithm for the Longest Tandem Scattered Subsequence Problem	93
<i>Adrian Kosowski</i>	

Automatic Document Categorization Based on k-NN and Object-Based Thesauri	101
<i>Sun Lee Bang, Hyung Jeong Yang, and Jae Dong Yang</i>	
Indexing Text Documents Based on Topic Identification	113
<i>Manonton Butarbutar and Susan McRoy</i>	
Cross-Comparison for Two-Dimensional Text Categorization	125
<i>Giorgio Maria Di Nunzio</i>	
DDOC: Overlapping Clustering of Words for Document Classification	127
<i>Guillaume Cleuziou, Lionel Martin, Viviane Clavier, and Christel Vrain</i>	
Evaluation of Web Page Representations by Content Through Clustering	129
<i>Arantza Casillas, Víctor Fresno, M. Teresa González de Lena, and Raquel Martínez</i>	
Evaluating Relevance Feedback and Display Strategies for Searching on Small Displays	131
<i>Vishwa Vinay, Ingemar J. Cox, Natasa Milic-Frayling, and Ken Wood</i>	
Information Extraction by Embedding HMM to the Set of Induced Linguistic Features	134
<i>Hyun Chul Lee and Jian Chang Mao</i>	
Finding Cross-Lingual Spelling Variants	136
<i>Krister Lindén</i>	
An Efficient Index Data Structure with the Capabilities of Suffix Trees and Suffix Arrays for Alphabets of Non-negligible Size	138
<i>Dong Kyue Kim, Jeong Eun Jeon, and Heejin Park</i>	
An Alphabet-Friendly FM-Index	150
<i>Paolo Ferragina, Giovanni Manzini, Veli Mäkinen, and Gonzalo Navarro</i>	
Concurrency Control and I/O-Optimality in Bulk Insertion	161
<i>Kerttu Pollari-Malmi and Eljas Soisalon-Soininen</i>	
Processing Conjunctive and Phrase Queries with the Set-Based Model	171
<i>Bruno Pôssas, Nivio Ziviani, Berthier Ribeiro-Neto, and Wagner Meira Jr.</i>	
Metric Indexing for the Vector Model in Text Retrieval	183
<i>Tomáš Skopal, Pavel Moravec, Jaroslav Pokorný, and Václav Snášel</i>	
Negations and Document Length in Logical Retrieval	196
<i>David E. Losada and Alvaro Barreiro</i>	

An Improvement and an Extension on the Hybrid Index for Approximate String Matching	208
<i>Heikki Hyvrö</i>	
First Huffman, Then Burrows-Wheeler: A Simple Alphabet-Independent FM-Index	210
<i>Szymon Grabowski, Veli Mäkinen, and Gonzalo Navarro</i>	
Metric Indexes for Approximate String Matching in a Dictionary	212
<i>Kimmo Fredriksson</i>	
Simple Implementation of String B-Trees	214
<i>Joong Chae Na and Kunsoo Park</i>	
Alphabet Permutation for Differentially Encoding Text	216
<i>Gad M. Landau, Ofer Levi, and Steven Skiena</i>	
A Space-Saving Linear-Time Algorithm for Grammar-Based Compression	218
<i>Hiroshi Sakamoto, Takuya Kida, and Shinichi Shimozono</i>	
Simple, Fast, and Efficient Natural Language Adaptive Compression	230
<i>Nieves R. Brisaboa, Antonio Fariña, Gonzalo Navarro, and José R. Paramá</i>	
Searching XML Documents Using Relevance Propagation	242
<i>Karen Sauvagnat, Mohand Boughanem, and Claude Christent</i>	
Dealing with Syntactic Variation Through a Locality-Based Approach	255
<i>Jesús Vilares and Miguel A. Alonso</i>	
Efficient Extraction of Structured Motifs Using Box-Links	267
<i>Alexandra M. Carvalho, Ana T. Freitas, Arlindo L. Oliveira, and Marie-France Sagot</i>	
Efficient Computation of Balancedness in Binary Sequence Generators	269
<i>Pedro García-Mochales and Amparo Fúster-Sabater</i>	
On Asymptotic Finite-State Error Repair	271
<i>Manuel Vilares, Juan Otero, and Jorge Graña</i>	
New Algorithms for Finding Monad Patterns in DNA Sequences	273
<i>Ravi Vijaya Satya and Amar Mukherjee</i>	
Motif Extraction from Weighted Sequences	286
<i>Costas S. Iliopoulos, Katerina Perdikuri, Evangelos Theodoridis, Athanasios Tsakalidis, and Kostas Tsichlas</i>	
Longest Motifs with a Functionally Equivalent Central Block	298
<i>Maxime Crochemore, Raffaele Giancarlo, and Marie-France Sagot</i>	

On the Transformation Distance Problem 310
Behshad Behzadi and Jean-Marc Steyaert

On Classification of Strings 321
Eljas Soisalon-Soininen and Tatu Ylönen

Author Index 331