

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Juha Kärkkäinen Jens Stoye (Eds.)

Combinatorial Pattern Matching

23rd Annual Symposium, CPM 2012
Helsinki, Finland, July 3-5, 2012
Proceedings

Volume Editors

Juha Kärkkäinen

University of Helsinki, Department of Computer Science

P.O. Box 68, Gustaf Hållströmin katu 2b, 00014 University of Helsinki, Finland

E-mail: juha.karkkainen@cs.helsinki.fi

Jens Stoye

Bielefeld University, Faculty of Technology

Universitätsstraße 25, 33615 Bielefeld, Germany

E-mail: jens.stoye@uni-bielefeld.de

ISSN 0302-9743

e-ISSN 1611-3349

ISBN 978-3-642-31264-9

e-ISBN 978-3-642-31265-6

DOI 10.1007/978-3-642-31265-6

Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012939778

CR Subject Classification (1998): I.5, F.2, G.1, G.2.1-2, G.2, E.1, E.4, H.3.3, I.2.7, H.2.8, J.3

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The papers contained in this volume were selected to be presented at the 23rd Annual Symposium on Combinatorial Pattern Matching (CPM 2012) held in Helsinki, Finland, during July 3–5, 2012, co-located with the 13th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2012).

The conference program included 33 contributed papers and two invited talks by Ron Shamir from Tel Aviv University and by Gonzalo Navarro from University of Chile in Santiago. The contributed papers were selected out of 60 submissions from 28 countries; each paper was reviewed by at least three reviewers. The joint program with the co-located SWAT 2012 conference included additional 34 contributed papers and two invited talks, by Joseph S.B. Mitchell from State University of New York at Stony Brook and by Roger Wattenhofer from ETH Zürich, published in a separate volume.

The Program Committee decided to grant the Best Student Paper Award to Pawel Gawrychowski for the paper titled “Simple and Efficient LZW-Compressed Multiple Pattern Matching”.

The objective of the annual CPM meetings is to provide an international forum for research in combinatorial pattern matching and related applications. It addresses issues of searching and matching strings and more complicated patterns such as trees, regular expressions, graphs, point sets, and arrays. The goal is to derive non-trivial combinatorial properties of such structures and to exploit these properties in order to either achieve superior performance for the corresponding computational problems or pinpoint conditions under which searches cannot be performed efficiently. The meeting also deals with problems in computational biology, data compression and data mining, coding, information retrieval, natural language processing, and pattern recognition.

The Annual Symposium on Combinatorial Pattern Matching started in 1990, and has since taken place every year. Previous CPM meetings were held in Paris, London (UK), Tucson, Padova, Asilomar, Helsinki, Laguna Beach, Aarhus, Piscataway, Warwick, Montreal, Jerusalem, Fukuoka, Morelia, Istanbul, Jeju Island, Barcelona, London (Canada), Ontario, Pisa, Lille, New York, and Palermo. Helsinki is the first place to host the meeting twice.

Starting from the third meeting, proceedings of all meetings have been published in the LNCS series, as volumes 644, 684, 807, 937, 1075, 1264, 1448, 1645, 1848, 2089, 2373, 2676, 3109, 3537, 4009, 4580, 5029, 5577, 6129, and 6661.

Selected papers from the first meeting appeared in volume 92 of *Theoretical Computer Science*, from the 11th meeting in volume 2 of the *Journal of Discrete Algorithms*, from the 12th meeting in volume 146 of *Discrete Applied Mathematics*, from the 14th meeting in volume 3 of the *Journal of Discrete Algorithms*, from the 15th meeting in volume 368 of *Theoretical Computer Science*, from the 16th meeting in volume 5 of the *Journal of Discrete Algorithms*, from the 19th

meeting in volume 410 of *Theoretical Computer Science*, from the 20th meeting in volume 9 of the *Journal of Discrete Algorithms*, from the 21st meeting in volume 213 of *Information and Computation*, and a special issue of *Theoretical Computer Science* is currently in preparation for selected papers from CPM 2011. A special issue of the *Journal of the Discrete Algorithms* is planned for the selected extended abstracts presented at this year's meeting.

We thank the CPM Steering Committee for supporting Helsinki as the site for CPM 2012, and for their advice and help in different issues. Financial and organizational support was provided by the University of Helsinki, Aalto University, and the Federation of the Finnish Learned Societies. The meeting would not have been possible without the laborious work of the local organizing teams of CPM 2012 and SWAT 2012.

The whole submission and review process was carried out with the invaluable help of the EasyChair conference system. Finally, special thanks are due to the members of the Program Committee and their subreviewers who worked very hard to ensure the timely review of all the submitted manuscripts, and participated in stimulating discussions that led to the selection of the papers for the conference.

April 2012

Juha Kärkkäinen
Jens Stoye

Organization

Program Committee

Hiroki Arimura	Hokkaido University, Japan
Diego Arroyuelo	Universidad Técnica Federico Santa María, Chile
Hideo Bannai	Kyushu University, Japan
Nieves R. Brisaboa	University of A Coruña, Spain
Ferdinando Cicalese	University of Salerno, Italy
Martin Farach-Colton	Rutgers University and Tokutek Inc., USA
Simone Faro	University of Catania, Italy
Kimmo Fredriksson	University of Eastern Finland, Finland
Inge Li Gørtz	Technical University of Denmark, Denmark
Shmuel Tomi Klein	Bar Ilan University, Israel
Roman Kolpakov	Moscow University, Russia
Juha Kärkkäinen	University of Helsinki, Finland (Co-chair)
Gad M. Landau	University of Haifa, Israel, and NYU-Poly, USA
Thierry Lecroq	University of Rouen, France
Bin Ma	University of Waterloo, Canada
Jan Manuch	University of British Columbia, Canada, and Simon Fraser University, Canada
Mehryar Mohri	Courant Institute of Mathematical Science, USA, and Google Research, New York, USA
Enno Ohlebusch	University of Ulm, Germany
Heejin Park	Hanyang University, South Korea
Mathieu Raffinot	LIAFA, Université Paris Diderot, France
Sven Rahmann	University of Duisburg-Essen, Germany
Wojciech Rytter	Warsaw University, Poland, and Copernicus University, Poland
Yasmín Ríos-Solís	Autonomous University of Nuevo León, Mexico
Marinella Sciortino	University of Palermo, Italy
Rahul Shah	Louisiana State University, USA
Jens Stoye	Bielefeld University, Germany (Co-chair)
Wing-Kin Sung	National University of Singapore, Singapore
Jorma Tarhio	Aalto University, Finland
German Tischler	The Wellcome Trust Sanger Institute, UK
Alexander Tiskin	University of Warwick, UK
Dekel Tsur	Ben Gurion University of the Negev, Israel

Steering Committee

Alberto Apostolico	University of Padova, Italy, and Georgia Institute of Technology, USA
Maxime Crochemore	Université Paris-Est, France, and King's, College London, UK
Zvi Galil	Georgia Institute of Technology, USA

Organizing Committee

Travis Gagie	Aalto University, Finland
Juha Kärkkäinen	University of Helsinki, Finland
Veli Mäkinen	University of Helsinki, Finland
Simon J. Puglisi	King's College London, UK
Leena Salmela	University of Helsinki, Finland
Jouni Sirén	University of Helsinki, Finland
Jorma Tarhio	Aalto University, Finland
Esko Ukkonen	University of Helsinki, Finland
Niko Välimäki	University of Helsinki, Finland

Additional Reviewers

Allauzen, Cyril	Hermelin, Danny
Amit, Mika	Hernández-Landa, Leonardo
Andonov, Rumen	I, Tomohiro
Bader, Martin	Ibarra-Rojas, Omar
Bansal, Mukul S.	Inenaga, Shunsuke
Bille, Philip	Kaltenbach, Hans-Michael
Blin, Guillaume	Kannan, Rajgopal
Breslauer, Dany	Karhu, Kalle
Castiglione, Giuseppa	Kim, Jin Wook
Cid-García, Néstor	Kim, Sung-Ryul
Cording, Patrick	Kociumaka, Tomasz
Czeizler, Elena	Kopelowitz, Tsvi
D'Addario, Marianna	Kowaluk, Mirosław
Dondi, Riccardo	Lee, Inbok
Ernst, Corinna	Lefebvre, Arnaud
Feng, Guangyu	Marschall, Tobias
Fertin, Guillaume	Mnich, Matthias
Gagie, Travis	Morales-Marroquín, Miguel
Gawrychowski, Paweł	Na, Joong Chae
Gog, Simon	Patil, Manish
Grabowski, Szymon	Prieur-Gaston, Élise
Groult, Richard	Puglisi, Simon J.
He, Lin	Riley, Michael

Rosone, Giovanna
 Rozenberg, Liat
 Salmela, Leena
 Sammeth, Michael
 Sim, Jeong Seop
 Sirén, Jouni
 Szreder, Bartosz
 Thachuk, Chris

Thankachan, Sharma V.
 Uno, Takeaki
 Venturini, Rossano
 Vialette, Stéphane
 Vildhøj, Hjalte Wedel
 Waleń, Tomasz
 Weimann, Oren
 Yang, Lian

Table of Contents

Invited Talks

Gene Regulation, Protein Networks and Disease: A Computational Perspective	1
<i>Ron Shamir</i>	
Wavelet Trees for All	2
<i>Gonzalo Navarro</i>	

Contributed Papers

The Maximum Number of Squares in a Tree	27
<i>Maxime Crochemore, Costas S. Iliopoulos, Tomasz Kociumaka, Marcin Kubica, Jakub Radoszewski, Wojciech Rytter, Wojciech Tyczyński, and Tomasz Waleń</i>	
Faster and Simpler Minimal Conflicting Set Identification (Extended Abstract)	41
<i>Aïda Ouangraoua and Mathieu Raffinot</i>	
Partitioning into Colorful Components by Minimum Edge Deletions	56
<i>Sharon Bruckner, Falk Hüffner, Christian Komusiewicz, Rolf Niedermeier, Sven Thiel, and Johannes Uhlmann</i>	
Approximation Algorithms and Hardness Results for Shortest Path Based Graph Orientations	70
<i>Dima Blokh, Danny Segev, and Roded Sharan</i>	
Constant-Time Word-Size String Matching	83
<i>Dany Breslauer, Leszek Gąsieniec, and Roberto Grossi</i>	
Pattern Matching in Multiple Streams	97
<i>Raphaël Clifford, Markus Jalsenius, Ely Porat, and Benjamin Sach</i>	
An Efficient Linear Pseudo-minimization Algorithm for Aho-Corasick Automata	110
<i>Omar AitMous, Frédérique Bassino, and Cyril Nicaud</i>	
Efficient Two-Dimensional Pattern Matching with Scaling and Rotation and Higher-Order Interpolation	124
<i>Christian Hundt and Florian Wendland</i>	

Hardness of Longest Common Subsequence for Sequences with Bounded Run-Lengths	138
<i>Guillaume Blin, Laurent Bulteau, Minghui Jiang, Pedro J. Tejada, and Stéphane Vialette</i>	
Near Linear Time Construction of an Approximate Index for All Maximum Consecutive Sub-sums of a Sequence	149
<i>Ferdinando Cicalese, Eduardo Laber, Oren Weimann, and Raphael Yuster</i>	
The Complexity of String Partitioning	159
<i>Anne Condon, Ján Maňuch, and Chris Thachuk</i>	
Towards an Optimal Space-and-Query-Time Index for Top- k Document Retrieval	173
<i>Wing-Kai Hon, Rahul Shah, and Sharma V. Thankachan</i>	
Document Listing for Queries with Excluded Pattern	185
<i>Wing-Kai Hon, Rahul Shah, Sharma V. Thankachan, and Jeffrey Scott Vitter</i>	
Cross-Document Pattern Matching	196
<i>Gregory Kucherov, Yakov Nekrich, and Tatiana Starikovskaya</i>	
FEMTO: Fast Search of Large Sequence Collections	208
<i>Michael P. Ferguson</i>	
Speeding Up q -Gram Mining on Grammar-Based Compressed Texts	220
<i>Keisuke Goto, Hideo Bannai, Shunsuke Inenaga, and Masayuki Takeda</i>	
Simple and Efficient LZW-Compressed Multiple Pattern Matching	232
<i>Paweł Gawrychowski</i>	
Computing the Burrows-Wheeler Transform of a String and Its Reverse	243
<i>Enno Ohlebusch, Timo Beller, and Mohamed I. Abouelhoda</i>	
Efficient Algorithm for Circular Burrows-Wheeler Transform	257
<i>Wing-Kai Hon, Tsung-Han Ku, Chen-Hua Lu, Rahul Shah, and Sharma V. Thankachan</i>	
Least Random Suffix/Prefix Matches in Output-Sensitive Time	269
<i>Niko Välimäki</i>	
Compressed String Dictionary Look-Up with Edit Distance One	280
<i>Djamal Belazzougui and Rossano Venturini</i>	
Time-Space Trade-Offs for Longest Common Extensions	293
<i>Philip Bille, Inge Li Gørtz, Benjamin Sach, and Hjalte Wedel Vildhøj</i>	

Local Exact Pattern Matching for Non-fixed RNA Structures	306
<i>Mika Amit, Rolf Backofen, Steffen Heyne, Gad M. Landau, Mathias Möhl, Christina Schmiedl, and Sebastian Will</i>	
Impact of the Energy Model on the Complexity of RNA Folding with Pseudoknots	321
<i>Saad Sheikh, Rolf Backofen, and Yann Ponty</i>	
Finding Longest Common Segments in Protein Structures in Nearly Linear Time	334
<i>Yen Kaow Ng, Hirotaka Ono, Ling Ge, and Shuai Cheng Li</i>	
A Linear Kernel for the Complementary Maximal Strip Recovery Problem	349
<i>Haitao Jiang and Binhai Zhu</i>	
Efficient Exponential Time Algorithms for Edit Distance between Unordered Trees	360
<i>Tatsuya Akutsu, Takeyuki Tamura, Daiji Fukagawa, and Atsuhiko Takasu</i>	
Fixed-Parameter Algorithms for Finding Agreement Supertrees	373
<i>David Fernández-Baca, Sylvain Guillemot, Brad Shuttters, and Sudheer Vakati</i>	
Computing the Rooted Triplet Distance between Galled Trees by Counting Triangles	385
<i>Jesper Jansson and Andrzej Lingas</i>	
Minimum Leaf Removal for Reconciliation: Complexity and Algorithms	399
<i>Riccardo Dondi and Nadia El-Mabrouk</i>	
On the Closest String via Rank Distance	413
<i>Liviu P. Dinu and Alexandru Popa</i>	
On Approximating String Selection Problems with Outliers	427
<i>Christina Boucher, Gad M. Landau, Avivit Levy, David Pritchard, and Oren Weimann</i>	
The Parameterized Complexity of the Shared Center Problem	439
<i>Zhi-Zhong Chen, Lusheng Wang, and Wenji Ma</i>	
Author Index	453