

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

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

University of Dortmund, Germany

Madhu Sudan

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

New York University, NY, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Alberto Apostolico Maxime Crochemore
Kunsoo Park (Eds.)

Combinatorial Pattern Matching

16th Annual Symposium, CPM 2005
Jeju Island, Korea, June 19-22, 2005
Proceedings

Volume Editors

Alberto Apostolico

University of Padova, Department of Information Engineering

Via Gradenigo 6a, 35131 Padova, Italy

E-mail: axa@dei.unipd.it

Maxime Crochemore

King's College London

and University of Marne-la-Vallée, Institut Gaspard-Monge

77454 Marne-la-Vallée, France

E-mail: maxime.crochemore@univ-mlv.fr

Kunsoo Park

Seoul National University

School of Computer Science and Engineering

Seoul 151-744, Korea

E-mail: kpark@snu.ac.kr

Library of Congress Control Number: 2005927143

CR Subject Classification (1998): F.2.2, I.5.4, I.5.0, I.7.3, H.3.3, J.3, E.4, G.2.1, E.1

ISSN 0302-9743

ISBN-10 3-540-26201-6 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-26201-5 Springer Berlin Heidelberg New York

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.

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik

Printed on acid-free paper SPIN: 11496656 06/3142 5 4 3 2 1 0

Preface

The 16th Annual Symposium on Combinatorial Pattern Matching was held on Jeju Island, Korea on June 19–22, 2005. Previous meetings were held in Paris, London, Tucson, Padova, Asilomar, Helsinki, Laguna Beach, Aarhus, Piscataway, Warwick, Montreal, Jerusalem, Fukuoka, Morelia, and Istanbul over the years 1990–2004.

In response to the call for papers, CPM 2005 received a record number of 129 papers. Each submission was reviewed by at least three Program Committee members with the assistance of external referees. Since there were many high-quality papers, the Program Committee’s task was extremely difficult. Through an extensive discussion the Program Committee accepted 37 of the submissions to be presented at the conference. They constitute original research contributions in combinatorial pattern matching and its applications.

In addition to the selected papers, CPM 2005 had three invited presentations, by Esko Ukkonen from the University of Helsinki, Ming Li from the University of Waterloo, and Naftali Tishby from The Hebrew University of Jerusalem.

We would like to thank all Program Committee members and external referees for their excellent work, especially given the demanding time constraints; they gave the conference its distinctive character. We also thank all who submitted papers for consideration; they all contributed to the high quality of the conference.

Finally, we thank the Organizing Committee members and the graduate students who worked hard to put in place the logistical arrangements of the conference. It is their dedicated contribution that made the conference possible and enjoyable.

June 2005

Alberto Apostolico, Maxime Crochemore, and Kunsoo Park

Organization

Program Committee

Tatsuya Akutsu	Kyoto University
<i>Alberto Apostolico</i>	University of Padova and Purdue University
Setsuo Arikawa	Kyushu University
<i>Maxime Crochemore</i>	University of Marne la Vallée
Martin Farach-Colton	Rutgers University
Costas S. Iliopoulos	King's College London
Juha Karkkainen	University of Helsinki
Dong Kyue Kim	Pusan National University
Tak-Wah Lam	University of Hong Kong
Moshe Lewenstein	Bar Ilan University
Bin Ma	University of Western Ontario
Giovanni Manzini	University of Piemonte Orientale
S. Muthukrishnan	Rutgers University
<i>Kunsoo Park</i>	Seoul National University
Mathieu Raffinot	CNRS
Rajeev Raman	University of Leicester
Cenk Sahinalp	Simon Fraser University
Andrew C. Yao	Tsinghua University
Frances F. Yao	City University of Hong Kong
Nivio Ziviani	Federal University of Minas Gerais

Steering Committee

Alberto Apostolico	University of Padova and Purdue University
Maxime Crochemore	University of Marne la Vallée
Zvi Galil	Columbia University

Organizing Committee

Dong Kyue Kim	Pusan National University
Yoo-Jin Chung	Hankuk University of Foreign Studies
Sung-Ryul Kim	Konkuk University
Heejin Park	Hanyang University
Jeong Seop Sim	Inha University
Jin Wook Kim	Seoul National University

Sponsoring Institutions

Ministry of Science and Technology, Korea
Seoul National University
SIGTCS of the Korea Information Science Society

External Referees

Julien Allali	
Marie-Pierre Béal	Tzvika Hartman
Anne Bergeron	Patrice Koehl
Marshall Bern	Tsvi Kopelowitz
Vincent Berry	Gregory Kucherov
Mathieu Blanchette	Thierry Lecroq
Dan Brown	Veli Mäkinen
Julien Clément	Wagner Meira, Jr.
Fabien Coulon	Nadia Pisanti
Thierson Couto Rosa	Sven Rahman
Fabiano Cupertino Botelho	Isidore Rigoutsos
Artur Czumaj	Eric Rivals
Davi de Castro Reis	Romeo Rizzi
Fabien de Montgolfier	Dominique Rossin
Funda Ergun	Kunihiko Sadakane
Paolo Ferragina	Jens Stoye
Guillaume Fertin	Jorma Tarhio
Dora Giammarresi	Stéphane Viallette
Sylvie Hamel	Hao-Chi Wong

Table of Contents

Sharper Upper and Lower Bounds for an Approximation Scheme for CONSENSUS-PATTERN	1
<i>Brona Brejová, Daniel G. Brown, Ian M. Harrower, Alejandro López-Ortiz, and Tomáš Vinař</i>	
On the Longest Common Rigid Subsequence Problem	11
<i>Bin Ma and Kaizhong Zhang</i>	
Text Indexing with Errors	21
<i>Moritz G. Maaß and Johannes Nowak</i>	
A New Compressed Suffix Tree Supporting Fast Search and Its Construction Algorithm Using Optimal Working Space	33
<i>Dong Kyue Kim and Heejin Park</i>	
Succinct Suffix Arrays Based on Run-Length Encoding	45
<i>Veli Mäkinen and Gonzalo Navarro</i>	
Linear-Time Construction of Compressed Suffix Arrays Using $o(n \log n)$ -Bit Working Space for Large Alphabets	57
<i>Joong Chae Na</i>	
Faster Algorithms for δ, γ -Matching and Related Problems	68
<i>Peter Clifford, Raphaël Clifford, and Costas Iliopoulos</i>	
A Fast Algorithm for Approximate String Matching on Gene Sequences ...	79
<i>Zheng Liu, Xin Chen, James Borneman, and Tao Jiang</i>	
Approximate Matching in the L_1 Metric	91
<i>Amihoud Amir, Ohad Lipsky, Ely Porat, and Julia Umanski</i>	
An Efficient Algorithm for Generating Super Condensed Neighborhoods ..	104
<i>Luís M.S. Russo and Arlindo L. Oliveira</i>	
The Median Problem for the Reversal Distance in Circular Bacterial Genomes	116
<i>Enno Ohlebusch, Mohamed Ibrahim Abouelhoda, Kathrin Hockel, and Jan Stalkamp</i>	
Using PQ Trees for Comparative Genomics	128
<i>Gad M. Landau, Lazmi Parida, and Oren Weimann</i>	
Hardness of Optimal Spaced Seed Design	144
<i>François Nicolas and Eric Rivals</i>	

Weighted Directed Word Graph	156
<i>Meng Zhang, Liang Hu, Qiang Li, and Jiubin Ju</i>	
Construction of Aho Corasick Automaton in Linear Time for Integer Alphabets	168
<i>Shiri Dori and Gad M. Landau</i>	
An Extension of the Burrows Wheeler Transform and Applications to Sequence Comparison and Data Compression	178
<i>Sabrina Mantaci, Antonio Restivo, G. Rosone, and Marinella Sciortino</i>	
DNA Compression Challenge Revisited: A Dynamic Programming Approach	190
<i>Behshad Behzadi and Fabrice Le Fessant</i>	
On the Complexity of Sparse Exon Assembly	201
<i>Carmel Kent, Gad M. Landau, and Michal Ziv-Ukelson</i>	
An Upper Bound on the Hardness of Exact Matrix Based Motif Discovery	219
<i>Paul Horton and Wataru Fujibuchi</i>	
Incremental Inference of Relational Motifs with a Degenerate Alphabet . . .	229
<i>Nadia Pisanti, Henry Soldano, and Mathilde Carpentier</i>	
Speeding up Parsing of Biological Context-Free Grammars	241
<i>Daniel Fredouille and Christopher H. Bryant</i>	
A New Periodicity Lemma	257
<i>Kangmin Fan, William F. Smyth, and R.J. Simpson</i>	
Two Dimensional Parameterized Matching	266
<i>Carmit Hazay, Moshe Lewenstein, and Dekel Tsur</i>	
An Optimal Algorithm for Online Square Detection	280
<i>Gen-Huey Chen, Jin-Ju Hong, and Hsueh-I Lu</i>	
A Simple Fast Hybrid Pattern-Matching Algorithm	288
<i>Frantisek Franek, Christopher G. Jennings, and William F. Smyth</i>	
Prefix-Free Regular-Expression Matching	298
<i>Yo-Sub Han, Yajun Wang, and Derick Wood</i>	
Reducing the Size of NFAs by Using Equivalences and Preorders	310
<i>Lucian Ilie, Roberto Solis-Oba, and Sheng Yu</i>	
Regular Expression Constrained Sequence Alignment	322
<i>Abdullah N. Arslan</i>	

A Linear Tree Edit Distance Algorithm for Similar Ordered Trees	334
<i>Hélène Touzet</i>	
A Polynomial Time Matching Algorithm of Ordered Tree Patterns Having Height-Constrained Variables	346
<i>Kazuhide Aikou, Yusuke Suzuki, Takayoshi Shoudai, Tomoyuki Uchida, and Tetsuhiro Miyahara</i>	
Assessing the Significance of Sets of Words	358
<i>Valentina Boeva, Julien Clément, Mireille Régnier, and Mathias Vandenbogaert</i>	
Inferring a Graph from Path Frequency	371
<i>Tatsuya Akutsu and Daiji Fukagawa</i>	
Exact and Approximation Algorithms for DNA Tag Set Design	383
<i>Ion I. Măndoiu and Dragoş Trincă</i>	
Parametric Analysis for Ungapped Markov Models of Evolution	394
<i>David Fernández-Baca and Balaji Venkatachalam</i>	
Linear Programming for Phylogenetic Reconstruction Based on Gene Rearrangements	406
<i>Jijun Tang and Bernard M.E. Moret</i>	
Identifying Similar Surface Patches on Proteins Using a Spin-Image Surface Representation	417
<i>Mary Ellen Bock, Guido M. Cortelazzo, Carlo Ferrari, and Concettina Guerra</i>	
Mass Spectra Alignments and Their Significance	429
<i>Sebastian Böcker and Hans-Michael Kaltenbach</i>	
Author Index	443