

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

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

Suleyman Cenk Sahinalp S. Muthukrishnan
Ugur Dogrusoz (Eds.)

Combinatorial Pattern Matching

15th Annual Symposium, CPM 2004
Istanbul, Turkey, July 5-7, 2004
Proceedings

Volume Editors

Suleyman Cenk Sahinalp
Simon Fraser University, School of Computing Science
Burnaby BC, V5A 1S6 Canada
E-mail: cenk@cs.sfu.ca

S. Muthukrishnan
Rutgers University, Department of Computer and Information Sciences
319 Core Bldg, 110 Frelinghuysen Rd, Piscataway, NJ 08854, USA
E-mail: muthu@cs.rutgers.edu

Ugur Dogrusoz
Bilkent University, Computer Engineering Department
06800 Ankara, Turkey
E-mail: ugur@cs.bilkent.edu.tr

Library of Congress Control Number: 2004108252

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

ISSN 0302-9743

ISBN 3-540-22341-X Springer-Verlag 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-Verlag. Violations are liable to prosecution under the German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11014331 06/3142 5 4 3 2 1 0

Preface

The 15th Annual Symposium on Combinatorial Pattern Matching was held in Ciragan Palace Hotel, Istanbul, Turkey during July 5–7, 2004. CPM 2004 repeated the success of its predecessors; it even surpassed them in terms of the number of invited speakers, the number of submissions, and the number of papers accepted and presented at the conference.

In response to the call for papers, CPM 2004 received a record number of 79 high-quality submissions. Each submission was reviewed by at least three program committee members and the comments were returned to the authors. Following an extensive electronic discussion period, the Program Committee accepted 36 of the submissions to be presented at the conference. They constitute original research contributions in combinatorial pattern matching algorithms and data structures, molecular sequence analysis, phylogenetic tree construction, and RNA and protein structure analysis and prediction.

CPM 2004 had five invited speakers. In alphabetical order they were: Evan Eichler from the University of Washington, USA, Martin Farach-Colton from Rutgers University, USA, Paolo Ferragina from the University of Pisa, Italy, Piotr Indyk from MIT, USA, and Gene Myers from the University of California, Berkeley, USA.

It is impossible to organize such a successful program without the help of many individuals. We would like to express our appreciation to the authors of the submitted papers and to the program committee members and external referees, who provided timely and significant reviews.

July 2004

S.C. Sahinalp,
S. Muthukrishnan,
U. Dogrusoz

Organization

CPM 2004 was locally organized by Bilkent University, Ankara, Turkey. Within Bilkent University, the Center for Bioinformatics (BCBI) and the Computer Engineering Department cooperated.



Executive Committee

| | |
|--------------------------|---|
| Organization Chair | Ugur Dogrusoz (Bilkent University) |
| Program Chairs | Suleyman Cenk Sahinalp (Simon Fraser University) |
| | S. Muthukrishnan (Rutgers University and AT&T) |
| Student Volunteer Chairs | Can Alkan (Simon Fraser University) |
| | Asli Ayaz (Bilkent University) |
| | Ozgun Babur (Bilkent University) |
| | Emek Demir (Bilkent University) |
| Social Events | Tasmanlar Tourism, Ankara, Turkey |
| Administrative Support | Gurkan Bebek (Case Western Reserve University) |
| | Emek Demir (Bilkent University) |

Program Committee

| | |
|------------------------|--|
| Gerth Stølting Brodal | University of Aarhus, Denmark |
| Jeremy Buhler | Washington University, USA |
| Ugur Dogrusoz | Bilkent University, Turkey |
| Zvi Galil | Columbia University, USA |
| Ramesh Hariharan | Indian Institute of Science, India |
| Ming Li | University of Waterloo, Canada |
| Stefano Lonardi | University of California, Riverside, USA |
| Ian Munro | University of Waterloo, Canada |
| Craig Neville Manning | Google Inc., USA |
| S. Muthukrishnan | Rutgers University and AT&T, USA |
| Joseph Nadeau | Case Western Reserve University, USA |
| Meral Ozsoyoglu | Case Western Reserve University, USA |
| Ely Porat | Bar-Ilan University, Israel |
| Kunihiko Sadakane | Kyushu University, Japan |
| Suleyman Cenk Sahinalp | Simon Fraser University, Canada |
| Mona Singh | Princeton University, USA |

Steering Committee

Alberto Apostolico

Purdue University, USA

and University of Padua, Italy

Maxime Crochemore

University of Marne la Vallée, France

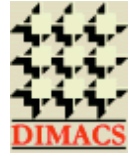
Zvi Galil

Columbia University, USA

Udi Manber

A9.com, USA

Sponsoring Organizations



External Referees

Can Alkan
 Abdullah Arslan
 Asli Ayaz
 Kensuke BaBa
 Ozgun Babur
 Gary Benson
 Petra Berenbrink
 Bernard Chazelle
 Tim Ting Chen
 Richard Cole
 Livio Colussi
 Maxime Crochemore
 Sergio De Agostino
 Emek Demir
 Zeynep Erson
 Kimmo Fredriksson
 Leszek Gasieniec
 Erhan Giral
 Danny Hermelin
 Lucian Ilie
 Tao Jiang
 Haim Kaplan
 Emre Karakoc
 Carmel Kent
 Ali Reza Khodabakhshi
 Takuya Kida
 Carl Kingsford
 Moshe Koppel
 Stefan Kurtz
 Jesper Larsson

Moshe Lewenstein
 Guohui Lin
 Hao Lin
 Bin Ma
 Veli Makinen
 Yishay Mansour
 Giovanni Manzini
 Gabriel Moruz
 Milan Mosny
 Giovanni Motta
 Giulio Pavesi
 Frederic Pio
 Teresa Przytycka
 Mathieu Raffinot
 Francesco Rizzo
 Igor B. Rogozin
 Wojciech Rytter
 Anoop Sarkar
 Nira Shafrir
 Dana Shapira
 Tetsuo Shibuya
 Dina Sokol
 Masayuki Takeda
 Andrey Utis
 Anil Vullikanti
 Oren Weimann
 Kaizhong Zhang
 Qiangfeng Zhang
 Zefeng Zhang
 Jie Zheng

Table of Contents

| | |
|--|-----|
| Sorting by Reversals in Subquadratic Time | 1 |
| <i>Eric Tannier and Marie-France Sagot</i> | |
| Computational Problems in Perfect Phylogeny Haplotyping: | |
| Xor-Genotypes and Tag SNPs | 14 |
| <i>Tamar Barzuza, Jacques S. Beckmann, Ron Shamir, and Itsik Pe'er</i> | |
| Sorting by Length-Weighted Reversals: Dealing with Signs and Circularity | 32 |
| <i>Firas Swidan, Michael A. Bender, Dongdong Ge, Simai He, Haodong Hu, and Ron Y. Pinter</i> | |
| Optimizing Multiple Spaced Seeds for Homology Search | 47 |
| <i>Jinbo Xu, Daniel G. Brown, Ming Li, and Bin Ma</i> | |
| Approximate Labelled Subtree Homeomorphism | 59 |
| <i>Ron Y. Pinter, Oleg Rokhlenko, Dekel Tsur, and Michal Ziv-Ukelson</i> | |
| On the Average Sequence Complexity | 74 |
| <i>Svante Janson, Stefano Lonardi, and Wojciech Szpankowski</i> | |
| Approximate Point Set Pattern Matching on Sequences and Planes | 89 |
| <i>Tomoaki Suga and Shinichi Shimozono</i> | |
| Finding Biclusters by Random Projections | 102 |
| <i>Stefano Lonardi, Wojciech Szpankowski, and Qiaofeng Yang</i> | |
| Real-Time String Matching in Sublinear Space | 117 |
| <i>Leszek Gąsieniec and Roman Kolpakov</i> | |
| On the k -Closest Substring and k -Consensus Pattern Problems | 130 |
| <i>Yishan Jiao, Jingyi Xu, and Ming Li</i> | |
| A Trie-Based Approach for Compacting Automata | 145 |
| <i>Maxime Crochemore, Chiara Epifanio, Roberto Grossi, and Filippo Mignosi</i> | |
| A Simple Optimal Representation for Balanced Parentheses | 159 |
| <i>Richard F. Geary, Naila Rahman, Rajeev Raman, and Venkatesh Raman</i> | |
| Two Algorithms for LCS Consecutive Suffix Alignment | 173 |
| <i>Gad M. Landau, Eugene Myers, and Michal Ziv-Ukelson</i> | |
| Efficient Algorithms for Finding Submasses in Weighted Strings | 194 |
| <i>Nikhil Bansal, Mark Cieliebak, and Zsuzsanna Lipták</i> | |

| | |
|---|-----|
| Maximum Agreement and Compatible Supertrees (Extended Abstract) . . . | 205 |
| <i>Vincent Berry and François Nicolas</i> | |
| Polynomial-Time Algorithms | |
| for the Ordered Maximum Agreement Subtree Problem | 220 |
| <i>Anders Dessmark, Jesper Jansson, Andrzej Lingas,</i> <i>and Eva-Marta Lundell</i> | |
| Small Phylogeny Problem: Character Evolution Trees | 230 |
| <i>Arvind Gupta, Ján Maňuch, Ladislav Stacho, and Chenchen Zhu</i> | |
| The Protein Sequence Design Problem in Canonical Model | |
| on 2D and 3D Lattices | 244 |
| <i>Piotr Berman, Bhaskar DasGupta, Dhruv Mubayi, Robert Sloan,</i> <i>György Turán, and Yi Zhang</i> | |
| A Computational Model for RNA Multiple Structural Alignment | 254 |
| <i>Eugene Davydov and Serafim Batzoglou</i> | |
| Computational Design of New and Recombinant Selenoproteins | 270 |
| <i>Rolf Backofen and Anke Busch</i> | |
| A Combinatorial Shape Matching Algorithm for Rigid Protein Docking . . | 285 |
| <i>Vicky Choi and Navin Goyal</i> | |
| Multi-seed Lossless Filtration (Extended Abstract) | 297 |
| <i>Gregory Kucherov, Laurent Noé, and Mikhail Roytberg</i> | |
| New Results for the 2-Interval Pattern Problem | 311 |
| <i>Guillaume Blin, Guillaume Fertin, and Stéphane Vialette</i> | |
| A Linear-Time Algorithm for Computing Translocation Distance | |
| between Signed Genomes | 323 |
| <i>Guojun Li, Xingqin Qi, Xiaoli Wang, and Binhai Zhu</i> | |
| Sparse Normalized Local Alignment | 333 |
| <i>Nadav Efraty and Gad M. Landau</i> | |
| Quadratic Time Algorithms for Finding Common Intervals | |
| in Two and More Sequences | 347 |
| <i>Thomas Schmidt and Jens Stoye</i> | |
| Maximal Common Connected Sets of Interval Graphs | 359 |
| <i>Michel Habib, Christophe Paul, and Mathieu Raffinot</i> | |
| Performing Local Similarity Searches with Variable Length Seeds | 373 |
| <i>Miklós Csűrös</i> | |
| Reversal Distance without Hurdles and Fortresses | 388 |
| <i>Anne Bergeron, Julia Mixtacki, and Jens Stoye</i> | |

| | |
|---|-----|
| A Fast Set Intersection Algorithm for Sorted Sequences | 400 |
| <i>Ricardo Baeza-Yates</i> | |
| Faster Two Dimensional Pattern Matching with Rotations | 409 |
| <i>Amihod Amir, Oren Kapah, and Dekel Tsur</i> | |
| Compressed Compact Suffix Arrays | 420 |
| <i>Veli Mäkinen and Gonzalo Navarro</i> | |
| Approximate String Matching Using Compressed Suffix Arrays | 434 |
| <i>Trinh N.D. Huynh, Wing-Kai Hon, Tak-Wah Lam,</i> <i>and Wing-Kin Sung</i> | |
| Compressed Index for a Dynamic Collection of Texts | 445 |
| <i>Ho-Leung Chan, Wing-Kai Hon, and Tak-Wah Lam</i> | |
| Improved Single and Multiple Approximate String Matching | 457 |
| <i>Kimmo Fredriksson and Gonzalo Navarro</i> | |
| Average-Case Analysis of Approximate Trie Search (Extended Abstract) | 472 |
| <i>Moritz G. Maaß</i> | |
| Author Index | 485 |