

Lecture Notes in Computer Science

1264

Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Advisory Board: W. Brauer D. Gries J. Stoer

Alberto Apostolico Jotun Hein (Eds.)

Combinatorial Pattern Matching

8th Annual Symposium, CPM 97
Aarhus, Denmark, June 30 - July 2, 1997
Proceedings



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany

Juris Hartmanis, Cornell University, NY, USA

Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Alberto Apostolico

Dipartimento di Elettronica e Informatica, University of Padova

Via Gradenigo 6/A, I-35131 Padova, Italy

E-mail: axa@artemide.dei.unipd.it

and

Department of Computer Sciences, Purdue University

West Lafayette, IN 47907, USA

E-mail: axa@cs.purdue.edu

Jotun Hein

University of Aarhus, Institute of Biological Science

Bldg. 540, Office 128

Ny Munkegade, DK-800 Aarhus C, Denmark

E-mail: jotun@pop.bio.aau.dk

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

**Combinatorial pattern matching : 8th annual symposium ;
proceedings / CPM 97, Aarhus, Denmark, June 30 - July 2, 1997.
Alberto Apostolico ; Jotun Hein (ed.). - Berlin ; Heidelberg ; New
York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ;
Santa Clara ; Singapore ; Tokyo : Springer, 1997
(Lecture notes in computer science ; Vol. 1264)
ISBN 3-540-63220-4**

CR Subject Classification (1991): F2.2, I.5.4, I.5.0, I.7.3, H.3.3, E.4, G.2.1,
J.3

ISSN 0302-9743

ISBN 3-540-63220-4 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 for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 1997

Printed in Germany

Typesetting: Camera-ready by author

SPIN 10548911 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Foreword

The papers contained in this volume were presented at the Eighth Annual Symposium on Combinatorial Pattern Matching (CPM 97), held June 30 – July 2, 1997, at the University of Aarhus, Denmark. They were selected from 32 papers submitted in response to a call for papers. In addition, invited lectures were given by A. Dress of the University of Bielefeld (Iterative versus Simultaneous Multiple Sequence Alignment) and J. B. Kruskal of Bell Labs (Modern Comparative Lexicostatistics).

Combinatorial Pattern Matching 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 for such structures and then to exploit these properties in order to achieve improved performance for the corresponding computational problems.

In recent years, a steady flow of high-quality research on this subject has changed a sparse set of isolated results into a full-fledged area of algorithmics with important applications. This area is expected to grow even further due to the increasing demand for speed and efficiency that comes especially from molecular biology, but also from areas such as information retrieval, pattern recognition, compiling, data compression, and program analysis. The objective of annual CPM gatherings is to provide an international forum for research in combinatorial pattern matching.

The general organization and orientation of CPM conferences is coordinated by a steering committee composed of A. Apostolico, M. Crochemore, Z. Galil, and U. Manber.

The first six meetings were held in Paris (1990), London (1991), Tucson (1992), Padova (1993), Pacific Grove (1994), Helsinki (1995), and Laguna Beach (1996). After the first meeting, a selection of the papers appeared as a special issue of *Theoretical Computer Science*. Since the third meeting, the proceedings have appeared as volumes 644, 684, 807, 937, and 1075 of the Lecture Notes in Computer Science series.

CPM 97 was organized by Jotun Hein of the Department of Genetics and Microbial Ecology at the University of Aarhus and Christian Nørgaard Storm Pedersen of the Department of Computer Science at the University of Aarhus. The conference was supported in part by the University of Aarhus, BRICS, and the National Research Foundation of Denmark.

Program Committee

A. Amir	S. Kannan
A. Apostolico <i>co-chair</i>	R. Kosaraju
R. Baeza-Yates	R. Shamir
D. Breslauer	S. Skyum
J. Hein <i>co-chair</i>	E. Ukkonen
D. Hirschberg	K. Zhang

Additional Referees

Vineet Bafna	Stefano Lonardi
Benny Chor	Rune Lyngsø
Livio Colussi	Dalit Naor
David Eppstein	Christian N. Storm P.
Roberto Grossi	Itsik Péér
Haim Kaplan	Theis Rauhe
Richard H. Lathrop	Haim Wolfson

Table of Contents

An Improved Pattern Matching Algorithm for Strings in Terms of Straight-Line Programs <i>Masamichi Miyazaki, Ayumi Shinohara, Masayuki Takeda</i>	1
Episode Matching <i>Gautam Das, Rudolf Fleischer, Leszek Gasieniec, Dimitris Gunopulos, Juha Kärkkäinen</i>	12
Efficient Algorithms for Approximate String Matching with Swaps <i>Jee-Soo Lee, Dong Kyue Kim, Kunsoo Park, Yookun Cho</i>	28
On the Complexity of Pattern Matching for Highly Compressed Two-Dimensional Texts <i>Piotr Berman, Marek Karpinski, Lawrence L. Larmore, Wojciech Plandowski, Wojciech Rytter</i>	40
Estimating the Probability of Approximate Matches <i>Stefan Kurtz, Gene Myers</i>	52
Space- and Time-Efficient Decoding with Canonical Huffman Trees <i>Shmuel T. Klein</i>	65
On Weak Circular Squares in Binary Words <i>Aviezri S. Fraenkel, Jamie Simpson, Mike Paterson</i>	76
An Easy Case of Sorting by Reversals <i>Nicholas Tran</i>	83
External Inverse Pattern Matching <i>Leszek Gasieniec, Piotr Indyk, Piotr Krysta</i>	90
Distributed Generation of Suffix Arrays <i>Gonzalo Navarro, João Paulo Kitajima, Berthier A. Ribeiro-Neto, Nivio Ziviani</i>	102
Direct Construction of Compact Directed Acyclic Word Graphs <i>Maxime Crochemore, Renaud Véroin</i>	116
Approximation Algorithms for the Fixed-Topology Phylogenetic Number Problem <i>Mary Cryan, Leslie Ann Goldberg, Cynthia A. Phillips</i>	130
A New Algorithm for the Ordered Tree Inclusion Problem <i>Thorsten Richter</i>	150

On Incremental Computation of Transitive Closure and Greedy Alignment <i>Saïd Abdeddaïm</i>	167
Aligning Coding DNA in the Presence of Frame-Shift Errors <i>Lars Arvestad</i>	180
A Filter Method for the Weighted Local Similarity Search Problem <i>Enno Ohlebusch</i>	191
Trie-Based Data Structures for Sequence Assembly <i>Ting Chen, Steven S. Skiena</i>	206
Flexible Identification of Structural Objects in Nucleic Acid Sequences: Palindromes, Mirror Repeats, Pseudoknots and Triple Helices <i>Marie-France Sagot, Alain Viari</i>	224
Banishing Bias from Consensus Sequences <i>Amir Ben-Dor, Guiseppe Lancia, Jennifer Perone, R. Ravi</i>	247
On the Nadeau-Taylor Theory of Conserved Chromosome Segments <i>David Sankoff, Marie-Noelle Parent, Isabelle Marchand, Vincent Ferretti</i>	262
Iterative versus Simultaneous Multiple Sequence Alignment <i>Andreas Dress</i>	275
Modern Comparative Lexicostatistics <i>Joseph B. Kruskal</i>	276
Author Index	277