

Lecture Notes in Computer Science

937

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

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

Zvi Galil Esko Ukkonen (Eds.)

Combinatorial Pattern Matching

6th Annual Symposium, CPM 95
Espoo, Finland, July 5-7, 1995
Proceedings



Springer

Series Editors

Gerhard Goos

Universität Karlsruhe

Vincenz-Priessnitz-Straße 3, D-76128 Karlsruhe, Germany

Juris Hartmanis

Department of Computer Science, Cornell University

4130 Upson Hall, Ithaca, NY 14853, USA

Jan van Leeuwen

Department of Computer Science, Utrecht University

Padualaan 14, 3584 CH Utrecht, The Netherlands

Volume Editors

Zvi Galil

Computer Science Department, Columbia University

New York, NY 10027, USA

Esko Ukkonen

Department of Computer Science, University of Helsinki

P.O. Box 26, FIN-00014, Helsinki, Finland

CR Subject Classification (1991): F.2.2, I.5.4, I.5.0, I.7.3, H.3.3, E.4, G.2.1, J.3

ISBN 3-540-60044-2 Springer-Verlag Berlin Heidelberg New York

CIP data applied for

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 1995

Printed in Germany

Typesetting: Camera-ready by author

SPIN: 10486313

06/3142-543210 - Printed on acid-free paper

Foreword

The papers contained in this volume were presented at the Sixth Annual Symposium on Combinatorial Pattern Matching (CPM 95), held July 5–7, 1995, at Hanasaari, a congress center located in Espoo, Finland, a few kilometers from Helsinki. They were selected from 44 papers submitted in response to the call of papers. In addition, invited lectures were given by Uzi Vishkin (On a Technique for Parsing a String) and H. W. Mewes (Genome Analysis: Pattern Search in Biological Sequences).

Combinatorial Pattern Matching addresses issues of searching and matching strings and more complicated patterns such as trees, regular expressions, extended expressions, etc. 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 the 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 five meetings were held in Paris (1990), London (1991), Tucson (1992), Padova (1993), and Pacific Grove (1994). 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, and 807 of the present series.

CPM 95 was organized by the Department of Computer Science at the University of Helsinki. The Local Organizing Committee of CPM 95 consisted of P. Kilpeläinen, J. Kärkkäinen, E. Sutinen, and J. Tarhio. The conference was supported in part by the Academy of Finland, Ministry of Education (Finland), and the University of Helsinki. The efforts of all are gratefully acknowledged.

Helsinki and New York, April 1995

Zvi Galil
Esko Ukkonen

Program Committee

A. Ehrenfeucht	P. Pevzner
Z. Galil, <i>co-chair</i>	I. Simon
D. Gusfield	J. Storer
U. Manber	E. Ukkonen, <i>co-chair</i>
M. Paterson	M. Wegman

Additional Referees

R. Agrawalla
D. Breslauer
M. Farach
P. Feofiloff
J. Field
R. Giancarlo
L. Guerra
M. Jain
P. Kilpeläinen
G. Kock
Y. Kohayakawa
A. Pereira do Lago

J. Meidanis
W. Miller
K. Murphy
K. Park
T. Przytycka
R. Ravi
J. Setubal
J. Soares
S. W. Song
P. Stelling
J. Tarhio
Y. Wakabayashi

Table of Contents

Computing Similarity between RNA Strings	1
<i>Vineet Bafna, S. Muthukrishnan, R. Ravi</i>	
Of Chicken Teeth and Mouse Eyes, or Generalized Character Compatibility	17
<i>Craig Benham, Sampath Kannan, Tandy Warnow</i>	
Efficient String Matching on Coded Texts	27
<i>Dany Breslauer, Leszek Gąsieniec</i>	
Fast Approximate Matching using Suffix Trees	41
<i>Archie L. Cobbs</i>	
Common Subsequences and Supersequences and Their Expected Length ...	55
<i>Vlado Dančik</i>	
Pattern Matching in Directed Graphs	64
<i>Jianghai Fu</i>	
Constant-Space String Matching with Smaller Number of Comparisons: Sequential Sampling	78
<i>Leszek Gąsieniec, Wojciech Plandowski, Wojciech Rytter</i>	
Multi-Dimensional Pattern Matching with Dimensional Wildcards	90
<i>Raffaele Giancarlo, Roberto Grossi</i>	
Minimizing Phylogenetic Number to Find Good Evolutionary Trees	102
<i>Leslie Ann Goldberg, Paul W. Goldberg, Cynthia A. Phillips, Elizabeth Sweedyk, Tandy Warnow</i>	
Making the Shortest-Paths Approach to Sum-of-Pairs Multiple Sequence Alignment More Space Efficient in Practice	128
<i>Sandeep K. Gupta, John D. Kececioglu, Alejandro A. Schäffer</i>	
An Efficient Algorithm for Developing Topologically Valid Matchings	144
<i>Liz Hanks, Ron K. Cytron, Will Gillett</i>	
Polynomial-Time Algorithm for Computing Translocation Distance between Genomes	162
<i>Sridhar Hannenhalli</i>	
On the Complexity of Comparing Evolutionary Trees	177
<i>Jotun Hein, Tao Jiang, Lusheng Wang, Kaizhong Zhang</i>	
Suffix Cactus: A Cross between Suffix Tree and Suffix Array	191
<i>Juha Kärkkäinen</i>	
Pattern-Matching for Strings with Short Descriptions	205
<i>Marek Karpinski, Wojciech Rytter, Ayumi Shinohara</i>	

Pairwise Alignment with Scoring on Tuples	215
<i>Lukas Knecht</i>	
Matching a Set of Strings with Variable Length Don't Cares	230
<i>Gregory Kucherov, Michaël Rusinowitch</i>	
Three-Dimensional Pattern Matching in Protein Structure Analysis	248
<i>Arthur M. Lesk</i>	
Genome Analysis: Pattern Search in Biological Macromolecules (Invited Lecture)	261
<i>H. W. Mewes, K. Heumann</i>	
Matching Patterns of an Automaton	286
<i>Mehryar Mohri</i>	
New Results and Open Problems Related to Non-Standard Stringology ...	298
<i>S. Muthukrishnan</i>	
String Matching in Hypertext	318
<i>Kunsoo Park, Dong Kyue Kim</i>	
Approximation Algorithms for Multiple Sequence Alignment under a Fixed Evolutionary Tree	330
<i>R. Ravi, John D. Kececioglu</i>	
A New Flexible Algorithm for the Longest Common Subsequence Problem	340
<i>Claus Rick</i>	
Smaller Representations for Finite-State Transducers and Finite-State Automata	352
<i>Emmanuel Roche</i>	
Multiple Sequence Comparison — A Peptide Matching Approach	366
<i>Marie-France Sagot, Alain Viari, Henri Soldano</i>	
On a Technique for Parsing a String (Invited Lecture)	386
<i>Uzi Vishkin</i>	
Dictionary Look-Up with Small Errors	387
<i>Andrew C. Yao, Frances F. Yao</i>	
On the Editing Distance between Undirected Acyclic Graphs and Related Problems	395
<i>Kaizhong Zhang, Jason T. L. Wang, Dennis Shasha</i>	
Author Index	409