

# Lecture Notes in Computer Science

684

Edited by G. Goos and J. Hartmanis

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



A. Apostolico M. Crochemore  
Z. Galil U. Manber (Eds.)

# Combinatorial Pattern Matching

4th Annual Symposium, CPM 93  
Padova, Italy, June 2-4, 1993  
Proceedings



**Springer-Verlag**

Berlin Heidelberg New York  
London Paris Tokyo  
Hong Kong Barcelona  
Budapest

Series Editors

Gerhard Goos  
Universität Karlsruhe  
Postfach 69 80  
Vincenz-Priessnitz-Straße 1  
W-7500 Karlsruhe, FRG

Juris Hartmanis  
Cornell University  
Department of Computer Science  
4130 Upson Hall  
Ithaca, NY 14853, USA

Volume Editors

Alberto Apostolico  
Computer Science Dept., Purdue University  
West Lafayette, IN 47907-1398, USA  
and  
Dipartimento di Elettronica e Informatica, Università di Padova  
I-35131 Padova, Italy

Maxime Crochemore  
L.I.T.P., Université Paris VII  
F-75251 Paris CX 05, France

Zvi Galil  
Columbia University, New York, NY 10027, USA  
and  
Tel Aviv University, Ramat Aviv, Tel Aviv, Israel

Udi Manber  
Computer Science Dept., University of Arizona  
Gould-Simpson 721. Tucson, AZ 85721, USA

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 3-540-56764-X Springer-Verlag Berlin Heidelberg New York  
ISBN 0-387-56764-X Springer-Verlag New York Berlin Heidelberg

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 1993  
Printed in Germany

Typesetting: Camera ready by author  
Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.  
45/3140-543210 - Printed on acid-free paper

# Foreword

The papers contained in this volume were presented at the fourth annual symposium on Combinatorial Pattern Matching, held June 2-4, 1993 in Padova, Italy. They were selected from 34 abstracts submitted in response to the call for papers.

Combinatorial Pattern Matching addresses issues of searching and matching of strings and more complicated patterns such as trees, regular expressions, extended expressions, etc. The goal is to derive nontrivial combinatorial properties for such structures and then to exploit these properties in order to achieve superior performances for the corresponding computational problems.

In recent years, a steady flow of high-quality scientific study of this subject has changed a sparse set of isolated results into a full-fledged area of algorithmics. This area is expected to grow even further due to the increasing demand for speed and efficiency that comes especially from molecular biology and the Genome project, but also from other diverse areas such as information retrieval (e.g., supporting complicated search queries), pattern recognition (e.g., using strings to represent polygons and string matching to identify them), compilers (e.g., using tree matching), data compression, and program analysis (e.g., program integration efforts). The stated objective of CPM gatherings is to bring together once a year the researchers active in the area for an informal and yet intensive exchange of information about current and future research in the area.

The first three meetings were held at the University of Paris in 1990, at the University of London in 1991, and at the University of Arizona, Tucson, in 1992. The first two meetings were informal and no proceedings were produced. The proceedings of the third meeting appeared as Volume 644 in this series.

Several external referees helped with the selection of papers for CPM 93. The Local Organizing Committee consisted of G. Bilardi, L. Colussi, C. Guerra and L. Toniolo. The Padova Ricerche Consortium provided services and, together with other sponsors, funds for the conference. The efforts of all are gratefully acknowledged. Special thanks are due to F. Bombi for eagerly promoting CPM 93 both within the University of Padova and outside of it.

March 1993

## Programm Committee

A. Apostolico, *chair*

M. Crochemore

A. Ehrenfeucht

A.S. Fraenkel

Z. Galil

G. Gonnet

D.S. Hirschberg

U. Manber

E.W. Myers

M.S. Waterman

## Table of Contents

A Linear Time Pattern Matching Algorithm Between a String and a Tree <i>Tatsuya Akutsu</i>	1
Tight Comparison Bounds for the String Prefix-Matching Problem <i>Dany Breslauer, Livio Colussi and Laura Toniolo</i>	11
3-D Docking of Protein Molecules <i>Daniel Fischer, Raquel Norel, Ruth Nussinov, and Haim J. Wolfson</i>	20
Minimal Separators of Two Words <i>Emmanuelle Garel</i>	35
Covering a String <i>Costas S. Iliopoulos, Dennis W.G. Moore and Kunsoo Park</i>	54
On the Worst-Case Behaviour of Some Approximation Algorithms for the Shortest Common Supersequence of $k$ Strings <i>Robert W. Irving and Campbell B. Fraser</i>	63
An Algorithm for Locating Non-Overlapping Regions of Maximum Alignment Score <i>Sampath K. Kannan and Eugene W. Myers</i>	74
Exact and Approximation Algorithms for the Inversion Distance Between Two Chromosomes <i>John Kececioğlu and David Sankoff</i>	87
The Maximum Weight Trace Problem in Multiple Sequence Alignment <i>John Kececioğlu</i>	106
An Algorithm for Approximate Tandem Repeats <i>Gad M. Landau and Jeanette P. Schmidt</i>	120
Two Dimensional Pattern Matching in a Digitized Image <i>Gad M. Landau and Uzi Vishkin</i>	134
Analysis of a String Edit Problem in a Probabilistic Framework <i>Guy Louchard and Wojciech Szpankowski</i>	152
Detecting False Matches in String Matching Algorithms <i>S. Muthukrishnan</i>	164

<b>On Suboptimal Alignments of Biological Sequences</b> <i>Dalit Naor and Douglas Brutlag</i>	179
<b>A Fast Filtration Algorithm for the Substring Matching Problem</b> <i>Pavel A. Pevzner and Michael S. Waterman</i>	197
<b>A Unifying Look at <math>d</math>-Dimensional Periodicities and Space Coverings</b> <i>Mireille Régnier and Ladan Rostami</i>	215
<b>Approximate String-Matching over Suffix Trees</b> <i>Esko Ukkonen</i>	228
<b>Multiple Sequence Comparison and <math>n</math>-Dimensional Image Reconstruction</b> <i>Martin Vingron and Pavel A. Pevzner</i>	243
<b>A New Editing Based Distance Between Unordered Labeled Trees</b> <i>Kaizhong Zhang</i>	254