Lecture Notes in Computer Science 807
Edited by G. Goos and J. Hartmanis

Maxime Crochemore Dan Gusfield (Eds.)

Combinatorial Pattern Matching

5th Annual Symposium, CPM 94
Asilomar, CA, USA, June 5-8, 1994
Proceedings

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Volume Editors

Maxime Crochemore
Institut Gaspard Monge, Université de Marne la Vallée
F-93160 Noisy le Grand, France

Dan Gusfield
Department of Computer Science, University of California
Davis, CA 95616-8692, 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 3-540-58094-8 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-58094-8 Springer-Verlag New York Berlin Heidelberg

CIP data applied for

Typesetting: Camera-ready by author
SPIN: 10131146 45/3140-543210 - Printed on acid-free paper

Foreword

The papers contained in this volume were presented at the fifth annual symposium on Combinatorial Pattern Matching, held June 5 - 8, 1994 at Asilomar, California. They were selected from 41 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 non-trivial 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 various genome projects, but also from areas such as information retrieval, pattern recognition, compiling, data compression, and program analysis. The stated objective of annual CPM gatherings is to bring together active researchers for an intensive exchange of information about current and future research in combinatorial pattern matching.

The general organisation and orientations of CPM Conferences are coordinated by a Steering Committee composed of A. Apostolico, M. Crochemore, Z. Galil, and U. Manber.

The first four meetings were held at the University of Paris (1990), at the University of London (1991), at the University of Arizona (Tucson, 1992), and at the University of Padova (1993). After the first meeting, a selection of the papers appeared as a special issue of *Theoretical Computer Science*. The proceedings of the third and fourth meetings appeared as volumes 644 and 684 of the present series.

External referees who helped with the selection of papers for CPM 94 are gratefully acknoledged.

Line Fonfrède and Jean-Louis Barrière (Institut Gaspard Monge, France) provided technical help during the selection process. Debbie Chadwick (Computer Science Department, U.C. Davis) helped with secretarial and administrative work. The conference was supported in part by the National Science Foundation and the University of California, Davis. The efforts of all are gratefully acknowledged.

Program Committee

Table of Contents

Session 1: Alignments

A Space Efficient Algorithm for Finding the Best Non-Overlapping Alignment Score 1
Gary Benson

The Parameterized Complexity of Sequence Alignment and Consensus 15
Hans Bodlaender, Rodney G. Downey, Michael R. Fellows, Harold T. Wareham

Computing all Suboptimal Alignments in Linear Space 31
Kun-Mao Chao

Approximation Algorithms for Multiple Sequence Alignment 43
Vineet Bafna, Eugene L. Lawler, Pavel A. Pevzner

A Context Dependent Method for Comparing Sequences 54
Xiaoqiu Huang

Fast Identification of Approximately Matching Substrings 64
Archie L. Cobbs

Alignment of Trees – an Alternative to Tree Edit 75
Tao Jiang, Lusheng Wang, Kaizhong Zhang

Session 2: Various Matchings

Parametric Recomputing in Alignment Graphs 87
Xiaoqiu Huang, Pavel A. Pevzner, Webb Miller

A Lossy Data Compression Based on String Matching: Preliminary Analysis and Suboptimal Algorithms 102
Tomasz Luczak, Wojciech Szpankowski

A Text Compression Scheme that Allows Fast Searching Directly in the Compressed File 113
Udi Manber

An Alphabet-Independent Optimal Parallel Search for Three Dimensional Pattern 125
Marek Karpinski, Wojciech Rytter

Unit Route Upper Bound for String-Matching on Hypercube 136
Laurent Lestrée

Session 3: Combinatorial aspects

Computation of Squares in a String 146
S. Rao Kosaraju

Minimization of Sequential Transducers 151
Mehryar Mohri

Shortest Common Superstrings for Strings of Random Letters 164
Kenneth S. Alexander

Maximal Common Subsequences and Minimal Common Supersequences 173
Robert W. Irving, Campbell B. Fraser

Dictionary-Matching on Unbounded Alphabets:
Uniform-Length Dictionaries 184
Dany Breslauer

Proximity Matching Using Fixed-Queries Trees 198
Ricardo Baeza-Yates, Walter Cunto, Udi Manber, Sun Wu

Query Primitives for Tree-Structured Data 213
Pekka Kilpeläinen, Heikki Mannila

Session 4: More Bio-Informatics

Multiple Matching of Parameterized Patterns 226
Ramana M. Idury, Alejandro A. Schäffer

Approximate String Matching with Don't Care Characters 240
Tatsuya Akutsu

Matching with Matrix Norm Minimization 250
Shouwen Tang, Kaizhong Zhang, Xiaolin Wu

Approximate String Matching and Local Similarity 259
William I. Chang, Thomas G. Marr

Polynomial-Time Algorithms for Computing Characteristic Strings 274
Minoru Ito, Kuniyasu Shimizu, Michio Nakanishi, Akihiro Hashimoto

Recent Methods for RNA Modeling
Using Stochastic Context-Free Grammars 289
Yasubumi Sakakibara, Michael Brown, Richard Hughey, I. Saira Mian, Kimmen Sjölander, Rebecca C. Underwood, David Haussler

Efficient Bounds for Oriented Chromosome Inversion Distance 307
John Kececioglu, David Sankoff

Index of Authors 326