Lecture Notes in Computer Science 6661

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Raffaele Giancarlo Giovanni Manzini (Eds.)

Combinatorial Pattern Matching

22nd Annual Symposium, CPM 2011
Palermo, Italy, June 27-29, 2011
Proceedings

Springer

Volume Editors

Raffaele Giancarlo
University of Palermo, Department of Mathematics
Via Archirafi 34, 90123 Palermo, Italy
E-mail: raffaele@math.unipa.it

Giovanni Manzini
University of 'Piemonte Orientale', Department of Computer Science
Viale T. Michel 11, 15121 Alessandria, Italy
E-mail: manzini@mfn.unipmn.it

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-21457-8 e-ISBN 978-3-642-21458-5
DOI 10.1007/978-3-642-21458-5
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011928675

CR Subject Classification (1998): F.2, I.5, H.3.3, J.3, I.4.2, E.4, G.2.1, E.1

LNCS Sublibrary: SL 1 – Theoretical Computer Science and General Issues

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The papers contained in this volume were presented at the 22nd Annual Symposium on Combinatorial Pattern Matching (CPM 2011) held in Mondello (Palermo), Italy, during June 27–29, 2011.

All the papers presented at the conference are original research contributions. We received 70 submissions from 20 countries; each paper was reviewed by at least three reviewers. The whole submission and review process was carried out with the invaluable help of the EasyChair conference system.

The committee decided to accept 36 papers. The program also included three invited talks by Nello Cristianini from the University of Bristol, UK, Gadi Landau from the University of Haifa, Israel, and Martin Vingron from the Max Planck Institute for Molecular Genetics, Berlin, Germany.

The objective of the annual CPM meetings is to provide an international forum for research in combinatorial pattern matching and related applications. It 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 of such structures and to exploit these properties in order to either achieve superior performance for the corresponding computational problems or pinpoint conditions under which searches cannot be performed efficiently. The meeting also deals with problems in computational biology, data compression and data mining, coding, information retrieval, natural language processing, and pattern recognition.

The Annual Symposium on Combinatorial Pattern Matching started in 1990, and has since taken place every year. Previous CPM meetings were held in Paris, London, Tucson, Padova, Asilomar, Helsinki, Laguna Beach, Aarhus, Piscataway, Warwick, Montreal, Jerusalem, Fukuoka, Morelia, Istanbul, Jeju Island, Barcelona, London, Ontario, Pisa, Lille, and New York.

Starting from the third meeting, proceedings of all meetings have been published in the LNCS series, volumes 644, 684, 807, 937, 1075, 1264, 1448, 1645, 1848, 2089, 2373, 2676, 3109, 3537, 4009, 4580, 5029, 5577, and 6129.

Selected papers from the first meeting appeared in volume 92 of *Theoretical Computer Science*, from the 11th meeting in volume 2 of *Journal of Discrete Algorithms*, from the 12th meeting in volume 146 of *Discrete Applied Mathematics*, from the 14th meeting in volume 3 of *Journal of Discrete Algorithms*, from the 15th meeting in volume 368 of *Theoretical Computer Science*, from the 16th meeting in volume 5 of *Journal of Discrete Algorithms*, from the 19th meeting in volume 410 of *Theoretical Computer Science*, and from the 20th meeting in volume 9 of *Journal of Discrete Algorithms.*

For this year, a special issue of *Theoretical Computer Science* is already planned for expanded versions of selected extended abstracts presented at the symposium.

Special thanks are due to the members of the Program Committee who worked very hard to ensure the timely review of all the submitted manuscripts, and participated in stimulating discussions that led to the selection of the papers for the conference.

April 2011

Raffaele Giancarlo
Giovanni Manzini

Best Student Paper Award

This year the Program Committee Co-chairs and the Local Organizing Committee sponsored a Best Student Paper Award. The award was reserved for papers authored solely by PhD students or by researchers in their first year of a Post-Doc assignment.

Among the 70 submissions received by the Program Committee, five of them were eligible for the award. The committee decided unanimously to assign the award to the paper:

Succincter Text Indexing with Wildcards

Chris Thachuk
Department of Computer Science,
University of British Columbia, Vancouver, Canada

We study the problem of indexing text with wildcard positions, motivated by the challenge of aligning sequencing data to large genomes that contain millions of single nucleotide polymorphisms (SNPs) —positions known to differ between individuals. SNPs modeled as wildcards can lead to more informed and biologically relevant alignments. We improve the space complexity of previous approaches by giving a succinct index requiring $(2 + o(1))n \log \sigma + O(n) + O(d \log n) + O(k \log k)$ bits for a text of length n over an alphabet of size σ containing d groups of k wildcards. The new index is particularly favorable for larger alphabets and comparable for smaller alphabets, such as DNA. A key to the space reduction is a result we give showing how any compressed suffix array can be supplemented with auxiliary data structures occupying $O(n) + O(d \log \frac{n}{d})$ bits to also support efficient dictionary matching queries. We present a new query algorithm for our wildcard index that greatly reduces the query working space to $O(dm + m \log n)$ bits, where m is the length of the query. We note that compared to previous results this reduces the working space by two orders of magnitude when aligning short read data to the human genome.

Organization

Program Committee

Alexandr Andoni	Microsoft Research SVC, USA
Mikhail Atallah	Purdue University, USA
Jérémy Barbay	University of Chile, Chile
Frédérique Bassino	Université Paris 13, France
Anne Bergeron	Université du Québec, Montreal, Quebec
Raphaël Clifford	University of Bristol, UK
Aldo de Luca	University of Naples “Federico II”, Italy
Chiara Epifanio	University of Palermo, Italy
Johannes Fischer	Karlsruhe Institute of Technology, Germany
Travis Gagie	Aalto University, Finland
Raffaele Giancarlo	University of Palermo, Italy (Co-chair)
Danny Hermelin	Max Planck Institute for Informatics, Germany
Wing-Kai Hon	National Tsing Hua University, Taiwan
Juha Kärkkäinen	University of Helsinki, Finland
Giosué Lo Bosco	University of Palermo, Italy
Stefano Lonardi	University of California Riverside, USA
Sabrina Mantaci	University of Palermo, Italy
Giovanni Manzini	University of Piemonte Or., Italy (Co-chair)
Burkhard Morgenstern	University of Göttingen, Germany
J. Ian Munro	University of Waterloo, Canada
Veli Mäkinen	University of Helsinki, Finland
Joong Chae Na	Sejong University, South Korea
Christian Pedersen	Aarhus University, Denmark
Wojciech Plandowski	University of Warsaw, Poland
Simon J. Puglisi	RMIT, Australia
Rajeev Raman	University of Leicester, UK
Mireille Régnier	INRIA-Saclay, France
Kunihiko Sadakane	National Institute of Informatics, Japan
David Sankoff	University of Ottawa, Canada
Giorgio Satta	University of Padova, Italy
Srinivasa Rao Satti	Seoul National University, South Korea
Roded Sharan	TelAviv University, Israel
William F. Smyth	McMaster University, Canada
Peter Stadler	University of Leipzig, Germany
Gabriel Valiente	Technical University of Catalonia, Spain
Susana Vinga	Inesc-ID, Portugal
Siu-Ming Yiu	University of Hong Kong, Hong Kong
Michal Ziv-Ukelson	Ben-Gurion University of the Negev, Israel

Steering Committee

Alberto Apostolico	University of Padova, Italy, and Georgia Institute of Technology, USA
Maxime Crochemore	Université Paris-Est, France, and King's College London, UK
Zvi Galil	Georgia Institute of Technology, USA

Organizing Committee

Chiara Epifanio	University of Palermo, Italy
Raffaele Giancarlo	University of Palermo, Italy
Giosué Lo Bosco	University of Palermo, Italy
Sabrina Mantaci	University of Palermo, Italy

Web and Publications Committee

Fabio Bellavia	University of Palermo, Italy
Alessio Langiu	University of Palermo, Italy
Carmen Lupascu	University of Palermo, Italy
Giovanna Rosone	University of Palermo, Italy
Luca Pinello	Harvard University, USA
Filippo Utro	IBM T.J. Watson Research Center, USA

External Referees

Atkins, Leon
Bankevich, Anton
Belazzougui, Djamal
Belcaid, Mahdi
Blin, Guillaume
Bouvel, Mathilde
Brejova, Bronislava
Bruckner, Sharon
Bucci, Michelangelo
Béal, Marie-Pierre
Camacho, Philippe
Canovas, Rodrigo
Carpi, Arturo
Claude, Francisco
Clément, Julien
Culpepper, Shane
Davoodi, Pooya
De Luca, Alessandro
Duma, Denisa
Dvorkin, Mikhail
Elberfeld, Michael
Elloumi, Mourad
Farzan, Arash
Flamm, Christoph
Francisco, Alexandre
Fraser, Robert
Gamzu, Iftah
Giambruno, Laura
Gog, Simon
Gotthilf, Zvi
Grabowski, Szymon
He, Meng
Jalsenius, Markus
Jansson, Jesper
Jiang, Shuai
Kaltenbach, Hans-Michael

Karhumaki, Juhani
Kim, Sung-Ryul
Kopelowitz, Tsvi
Kreft, Sebastian
Kubica, Marcin
Kufleitner, Manfred
Kulikov, Alexander
Kuruppu, Shanika
Lee, Inbok
Liptak, Zsuzsanna
Ma, Jian
Mnich, Matthias
Mosig, Axel
Mozes, Shay
Nekrich, Yakov
Nicaud, Cyril
Nielsen, Jesper
Nikolenko, Sergey
Noé, Laurent
Parida, Laxmi
Park, Heejin
Pinello, Luca
Pinhas, Tamar
Polishko, Anton
Ponty, Yann
Popa, Alexandru
Poulalhon, Dominique
Radoszewski, Jakub
Restivo, Antonio
Russo, Luis
Rytter, Wojciech
Sach, Benjamin
Salmela, Leena
Sand, Andreas
Sanders, Peter
Schmiedl, Christina
Sciortino, Marinella
Segev, Danny
Silverbush, Dana
Sim, Jeong Seop
Simonsen, Martin
Simpson, Jamie
Sirotkin, Alexander
Sirén, Jouni
Speck, Jochen
Tarhio, Jorma
Tataru, Paula
Thankachan, Sharma V.
Toivonen, Jarkko
Ustaoglu, Berkant
Vaglica, Roberto
Valenzuela, Daniel
Verbin, Elad
Verzotto, Davide
Vialette, Stéphane
Vyahhi, Nikolay
Välimäki, Niko
Walen, Tomasz
Zakov, Shay
Zaroda, Artur
Zizza, Rosalba

Sponsoring Institutions

University of Palermo
University of Piemonte Orientale

Table of Contents

Invited Talks

Contributed Papers