LIPIcs – Leibniz International Proceedings in Informatics

LIPIcs is a series of high-quality conference proceedings across all fields in informatics. LIPIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

Editorial Board

- Luca Aceto (Chair, Gran Sasso Science Institute and Reykjavik University)
- Christel Baier (TU Dresden)
- Mikolaj Bojanczyk (University of Warsaw)
- Roberto Di Cosmo (INRIA and University Paris Diderot)
- Javier Esparza (TU München)
- Meena Mahajan (Institute of Mathematical Sciences)
- Dieter van Melkebeek (University of Wisconsin-Madison)
- Anca Muscholl (University Bordeaux)
- Luke Ong (University of Oxford)
- Catuscia Palamidessi (INRIA)
- Thomas Schwentick (TU Dortmund)
- Raimund Seidel (Saarland University and Schloss Dagstuhl – Leibniz-Zentrum für Informatik)

ISSN 1868-8969

https://www.dagstuhl.de/lipics
To all algorithmic stringologists in the world
Contents

Preface

Nadia Pisanti and Solon P. Pissis ........................................ 0:xi

Program Committee

.................................................................................. 0:xiii

External Reviewers

.................................................................................. 0:xv

Authors of Selected Papers

.................................................................................. 0:xvii

Invited Talks

How to Exploit Periodicity

Paweł Gawrychowski .................................................. 1:1–1:1

Some Variations on Lyndon Words

Francesco Dolce, Antonio Restivo, and Christophe Reutenauer ........ 2:1–2:14

Stringology Combats Microbiological Threats

Michal Ziv-Ukelson .................................................. 3:1–3:1

Regular Papers

Optimal Rank and Select Queries on Dictionary-Compressed Text

Nicola Prezza .......................................................... 4:1–4:12

A 2-Approximation Algorithm for the Complementary Maximal Strip Recovery Problem

Haitao Jiang, Jiong Guo, Daming Zhu, and Binhai Zhu .................... 5:1–5:13

Sufficient Conditions for Efficient Indexing Under Different Matchings

Amihood Amir and Eitan Kondratovsky .................................. 6:1–6:12

Space-Efficient Computation of the LCP Array from the Burrows-Wheeler Transform

Nicola Prezza and Giovanna Rosone .................................. 7:1–7:18

Safe and Complete Algorithms for Dynamic Programming Problems, with an Application to RNA Folding

Niko Kiirala, Leena Salmela, and Alexandru I. Tomescu .................. 8:1–8:16

Conversion from RLBWT to LZ77

Takaaki Nishimoto and Yasuo Tabei .................................. 9:1–9:12

Fully-Functional Bidirectional Burrows-Wheeler Indexes and Infinite-Order De Bruijn Graphs

Djamal Belazzougui and Fabio Cunial ................................ 10:1–10:15
### Contents

Entropy Lower Bounds for Dictionary Compression  
*Michał Gańczorz* ................................................................. 11:1–11:18

A New Class of Searchable and Provably Highly Compressible String Transformations  
*Raffaele Giancarlo, Giovanni Manzini, Giovanna Rosone, and Marinella Sciortino* 12:1–12:12

Compressed Multiple Pattern Matching  
*Dmitry Kosolobov and Nikita Sivukhin* ................................. 13:1–13:14

Hamming Distance Completeness  
*Karim Labib, Przemysław Uznański, and Daniel Wolleb-Graf* ........ 14:1–14:17

Approximating Approximate Pattern Matching  
*Jan Studený and Przemysław Uznański* ................................. 15:1–15:13

Cartesian Tree Matching and Indexing  

Indexing the Bijective BWT  
*Hideo Bannai, Juha Kärkkäinen, Dominik Köppl, and Marcin Piątkowski* 17:1–17:14

On Maximal Repeats in Compressed Strings  
*Julian Pape-Lange* ................................................................. 18:1–18:13

Dichotomic Selection on Words: A Probabilistic Analysis  
*Ali Akhavi, Julien Clément, Dimitri Darthenay, Loïck Lhote, and Brigitte Vallée* . 19:1–19:19

Finding a Small Number of Colourful Components  
*Laurent Bulteau, Konrad K. Dabrowski, Guillaume Fertin, Matthew Johnson,  
Daniel Paulusma, and Stéphane Vialette* ................................. 20:1–20:14

Streaming Dictionary Matching with Mismatches  
*Paweł Gawrychowski and Tatiana Starikovskaya* ..................... 21:1–21:15

Quasi-Periodicity in Streams  

Computing Runs on a Trie  
*Ryo Sugahara, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and  
Masayuki Takeda* ................................................................. 23:1–23:11

Linking BWT and XBW via Aho-Corasick Automaton: Applications to Run-Length Encoding  
*Bastien Cazaux and Eric Rivals* ............................................. 24:1–24:20

Quasi-Linear-Time Algorithm for Longest Common Circular Factor  
*Mai Alzamel, Maxime Crochemore, Costas S. Iliopoulos, Tomasz Kociumaka,  
Jakub Radoszewski, Wojciech Rytter, Juliusz Straszynski, Tomasz Wałen, and  
Wiktory Żuba* ........................................................................ 25:1–25:14

Simulating the DNA Overlap Graph in Succinct Space  
*Diego Díaz-Domínguez, Travis Gagie, and Gonzalo Navarro* .......... 26:1–26:20

Faster Queries for Longest Substring Palindrome After Block Edit  
*Mitsuru Funakoshi, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and  
Masayuki Takeda* .................................................................... 27:1–27:13
A Rearrangement Distance for Fully-Labelled Trees
   Giulia Bernardini, Paola Bonizzoni, Gianluca Della Vedova, and Murray Patterson 28:1–28:15

On the Size of Overlapping Lempel-Ziv and Lyndon Factorizations
   Yuki Urabe, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, and Masayuki Takeda ................................................................. 29:1–29:11

Online Algorithms for Constructing Linear-Size Suffix Trie
   Diptarama Hendrian, Takuya Takagi, and Shunsuke Inenaga 30:1–30:19

Searching Long Repeats in Streams
   Oleg Merkurev and Arseny M. Shur .................................................. 31:1–31:14

Computing the Antiperiod(s) of a String
The Annual Symposium on Combinatorial Pattern Matching (CPM) is the international research forum in the areas of combinatorial pattern matching, string algorithms and related applications. The studied objects include strings as well as trees, regular expressions, graphs, and point sets, and the goal is to design efficient algorithms and data structures based on their properties, in order to design efficient algorithmic solutions for the addressed computational problems. The problems this conference deals with include those in bioinformatics and computational biology, coding and data compression, combinatorics on words, data mining, information retrieval, natural language processing, pattern matching and discovery, string algorithms, string processing in databases, symbolic computation, and text searching and indexing. This volume contains the papers presented at the 30th Annual Symposium on Combinatorial Pattern Matching (CPM 2019) held on June 18-20, 2019 in Pisa, Italy. The conference programme includes 29 contributed papers and three invited talks by Paweł Gawrychowski (University of Wrocław, Poland), Antonio Restivo (University of Palermo, Italy), and Michal Ziv-Ukelson (Ben Gurion University of the Negev, Israel). Contributions of the invited lectures are also included in this volume. For the first time, this edition of CPM will also include the "Highlights of CPM" special session, for presenting the highlights of recent developments in combinatorial pattern matching. In this first edition we have invited Diptarka Chakraborty (Weizmann Institute of Science, Israel) to present his FOCS 2018 (best) paper "Approximating edit distance within constant factor in truly sub-quadratic time", and Nicola Prezza (University of Pisa, Italy) to present his STOC 2018 paper "At the roots of dictionary compression: string attractors". The contributed papers were selected out of 50 submissions, corresponding to an acceptance ratio of about 58%. Each submission received at least three reviews. We thank the members of the Programme Committee and all the additional external reviewers that are listed below for their hard and invaluable and collaborative work that resulted in an excellent scientific programme.

The Annual Symposium on Combinatorial Pattern Matching started in 1990, and has since then taken place every year. Previous CPM meetings were held in Paris, London (UK), Tucson, Padova, Asilomar, Helsinki, Laguna Beach, Aarhus, Piscataway, Warwick, Montreal, Jerusalem, Fukuoka, Morelia, Istanbul, Jeju Island, Barcelona, London (Ontario, Canada), Pisa, Lille, New York, Palermo, Helsinki, Bad Herrenalb, Moscow, Ischia, Tel Aviv, Warsaw, and Qingdao. From 1992 to the 2015 meeting, all proceedings were published in the LNCS (Lecture Notes in Computer Science) series. Since 2016, the CPM proceedings appear in the LIPIcs (Leibniz International Proceedings in Informatics) series, as volume 54 (CPM 2016), 78 (CPM 2017), and 105 (CPM 2018). The whole submission and review process was carried out with the help of the EasyChair conference system. We thank the CPM Steering Committee for supporting Pisa as the site for CPM 2019 and for their advice. The conference is generously hosted by the Computer Science Department of the University of Pisa. We thank Alessio Conte, Roberto Grossi, Veronica Guerrini, Giulio Pibiri, Nicola Prezza, Giovanna Rosone, Rossano Venturini (University of Pisa), and Andrea Marino (University of Florence) for their involvement and help in local arrangements. Finally, we would like to thank the MSCA RISE 2015 BIRDS Project “Bioinformatics and Information Retrieval Data Structures Analysis & Design” (GA 690941) and the UniPi PRA2017-44 Project for providing generous financial support to the conference.

Nadia Pisanti & Solon P. Pissis
Program Committee

Nadia Pisanti (Co-Chair)  
University of Pisa, Italy

Solon P. Pissis (Co-Chair)  
CWI Amsterdam, the Netherlands

Golnaz Badkobeh  
Goldsmiths University of London, UK

Carl Barton  
European Bioinformatics Institute, UK

Djamal Belazzougui  
DTISI-CERIST, Algeria

Christina Boucher  
University of Florida, USA

Gianluca Della Vedova  
University of Milano-Bicocca, Italy

Gabriele Fici  
University of Palermo Palermo, Italy

Szymon Grabowski  
Lodz University of Technology, Poland

Inge Li Gørtz  
Technical University of Denmark, Denmark

Wing-Kai Hon  
National Tsing Hua University, Taiwan

Tomohiro I  
Kyushu Institute of Technology, Japan

Tomasz Kociumaka  
University of Warsaw, Poland

Christian Komusiewicz  
Philipps-Universität Marburg, Germany

Tsvi Kopelowitz  
Bar Ilan University, Israel

Florin Manea  
Universität Kiel, Germany

Veli Mäkinen  
University of Helsinki, Finland

Robert Mercas  
Loughborough University, UK

Gonzalo Navarro  
University of Chile, Chile

Yakov Nekrich  
University of Waterloo, Canada

Cyril Nicaud  
Université Paris-Est, France

Alberto Policriti  
Università degli Studi di Udine, Italy

Simon J. Puglisi  
University of Helsinki, Finland

Jakub Radoszewski  
University of Warsaw, Poland

Giovanna Rosone  
University of Pisa, Italy

Eva Rotenberg  
Technical University of Denmark, Denmark

Marie-France Sagot  
INRIA and Univ. Claude Bernard, France

Jamie Simpson  
Curtin University, Australia

Tatiana Starikovskaya  
École Normale Supérieure, France

Jens Stoye  
Universität Bielefeld, Germany

Wing-Kin Sung  
National University of Singapore, Singapore

Sharma V. Thankachan  
University of Central Florida, USA

Oren Weimann  
University of Haifa, Israel

30th Annual Symposium on Combinatorial Pattern Matching (CPM 2019).
Editors: Nadia Pisanti and Solon P. Pissis
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
External Reviewers

Anders Aamand
Hideo Bannai
Frederique Bassino
Olivier Bodini
Tiziana Calamoneri
Bastien Cazaux
Panagiotis Charalampopoulos
Anders Roy Christiansen
Julien Clément
Anne Condon
Graham Cormode
Maxime Crochemore
Alessandro De Luca
Sebastian Deorowicz
Daniel Doerr
Guillaume Ducoffe
Bartłomiej Dudek
Pawel Gawrychowski
Simon Gog
Shay Golan
Niels Grüttemeier
Danny Hermelin
Matthew Hoban
Stepan Holub
Oded Lachish
Dominik Kempa
Christian Konrad
Dominik Köppl

Dmitry Kosolobov
Tomasz Marek Kowalski
Gregory Kucherov
Alan Kuhnle
Thierry Lecroq
Avivit Levy
Sabrina Mantaci
Daniel Martin
Yuto Nakashima
Pascal Ochem
Murray Patterson
Yann Ponty
Marco Previtali
Nicola Prezza
Eric Rivals
Raffaella Rizzi
Marinella Sciortino
Arseny Shur
Blerina Sinaimeri
Frank Sommer
Teresa Anna Steiner
Juliusz Straszysiński
Alexander Tiskin
Rossano Venturini
Tomasz Waleń
Roland Wittler
Robert Zimmer
Michal Ziv-Ukelson

30th Annual Symposium on Combinatorial Pattern Matching (CPM 2019).
Editors: Nadia Pisanti and Solon P. Pissis
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
Authors of Selected Papers

Mai Alzamel
Ali Akhavi
Hayam Alamro
Amir Amihood
Golnaz Badkobeh
Hideo Bannai
Djamal Belazzougui
Giulia Bernardini
Paola Bonizzoni
Laurent Bulteau
Bastien Cazaux
Julien Clément
Maxime Crochemore
Fabio Cunial
Konrad K. Dabrowski
Dimitri Darthenay
Gianluca Della Vedova
Diego Diaz-Dominguez
Guillaume Fertin
Mitsuru Funakoshi
Travis Gagie
Michal Gańczorz
Pawel Gawrychowski
Raffaele Giancarlo
Jiong Guo
Diptarama Hendrian
Costas S. Iliopoulos
Shunsuke Inenaga
Haitao Jiang
Matthew Johnson
Juha Kärkkäinen
Niko Kiirrala
Tomasz Kociumaka
Eitan Kondratovsky
Dominik Köppl
Dmitry Kosolobov
Karim Labib
Loïck Lhote
Gad M. Landau
Giovanni Manzini
Oleg Merkurev
Yuto Nakashima
Gonzalo Navarro
Takaaki Nishimoto
Julian Pape-Lange
Kunsoo Park
Sung Gwan Park
Murray Patterson
Daniel Paulusma
Marcin Piątkowski
Nicola Prezza
Simon J. Puglisi
Leena Salmela
Arseny Shur
Nikita Sivukhin
Tatiana Starikovskaya
Juliusz Straszęński
Jakub Radoszewski
Antonio Restivo
Eric Rivals
Giovanna Rosone
Wojciech Rytter
Marinella Sciortino
Jan Studený
Ryo Sugahara
Yasuho Tabei
Takuya Takagi
Masayuki Takeda
Alexandru I. Tomescu
Yuki Urabe
Przemysław Uznański
Brigitte Vallee
Stephane Vialette
Tomasz Waleń
Daniel Wolle-Gruf
Binhai Zhu
Daming Zhu
Michal Ziv-Ukelson
Wiktor Zuba