31st Annual Symposium on Combinatorial Pattern Matching

CPM 2020, June 17–19, 2020, Copenhagen, Denmark

Edited by

Inge Li Gørtz

Oren Weimann

Editors

Inge Li Gørtz
Technical University of Denmark, DTU Compute, Lyngby, Denmark
inge@dtu.dk

Oren Weimann
University of Haifa, Israel
oren@cs.haifa.ac.il

ACM Classification 2012
Theory of computation → Design and analysis of algorithms; Theory of computation → Pattern matching; Mathematics of computing → Discrete mathematics; Mathematics of computing → Information theory; Mathematics of computing → Combinatoric problems; Information systems → Information retrieval; Applied computing → Computational biology

ISBN 978-3-95977-149-8

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at https://www.dagstuhl.de/dagpub/978-3-95977-149-8.

Publication date
June, 2020

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at https://portal.dnb.de.

License
This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0):
https://creativecommons.org/licenses/by/3.0/legalcode.
In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:
- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/LIPIcs.CPM.2020.0

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

Contents

Preface
Inge Li Gørtz and Oren Weimann 0:ix

Program Committee
..... 0:xi

Reviewers
..... 0:xiii

Authors
..... 0:xv–0:xvi

Invited Talk

Algebraic Algorithms for Finding Patterns in Graphs
Thore Husfeldt 1:1–1:1

Regular Papers

Finding the Anticover of a String
Mai Alzamel, Alessio Conte, Shuhei Denzumi, Roberto Grossi, Costas S. Iliopoulos, Kazuhiro Kurita, and Kunihiro Wasa 2:1–2:11

Double String Tandem Repeats
Amihood Amir, Ayelet Butman, Gad M. Landau, Shoshana Marcus, and Dina Sokol 3:1–3:13

Efficient Tree-Structured Categorical Retrieval
Djamal Belazzougui and Gregory Kucherov 4:1–4:11

Time-Space Tradeoffs for Finding a Long Common Substring
Stav Ben-Nun, Shay Golan, Tomasz Kociumaka, and Matan Kraus 5:1–5:14

On Two Measures of Distance Between Fully-Labelled Trees
Giulia Bernardini, Paola Bonizzoni, and Paweł Gawrychowski 6:1–6:16

String Sanitization Under Edit Distance
Giulia Bernardini, Huiping Chen, Grigorios Loukides, Nadia Pisanti, Solon P. Pissis, Leen Stougie, and Michelle Sweering 7:1–7:14

Counting Distinct Patterns in Internal Dictionary Matching
Panagiotis Charalampopoulos, Tomasz Kociumaka, Manal Mohamed, Jakub Radoszewski, Wojciech Rytter, Juliusz Straszyński, Tomasz Waleń, and Wiktor Zuba 8:1–8:15

Dynamic String Alignment
Panagiotis Charalampopoulos, Tomasz Kociumaka, and Shay Mozes 9:1–9:13

Unary Words Have the Smallest Levenshtein k-Neighbourhoods
Panagiotis Charalampopoulos, Solon P. Pissis, Jakub Radoszewski, Tomasz Waleń, and Wiktor Zuba 10:1–10:12

31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020).
Editors: Inge Li Gørtz and Oren Weimann

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Summarizing Diverging String Sequences, with Applications to Chain-Letter Petitions
Patty Commins, David Liben-Nowell, Tina Liu, and Kiran Tomlinson 11:1–11:15

Detecting k-(Sub-)Cadences and Equidistant Subsequence Occurrences
Mitsuru Funakoshi, Yuto Nakashima, Shunsuke Inenaga, Hideo Bannai, Masayuki Takeda, and Ayumi Shinohara 12:1–12:11

FM-Index Reveals the Reverse Suffix Array
Arnab Ganguly, Daniel Gibney, Sahar Hooshmand, M. Oğuzhan Külekci, and Sharma V. Thankachan 13:1–13:14

On Indeterminate Strings Matching
Paweł Gawrychowski, Samah Ghazawi, and Gad M. Landau 14:1–14:14

The Streaming k-Mismatch Problem: Tradeoffs Between Space and Total Time
Shay Golan, Tomasz Kociumaka, Tsvi Kopelowitz, and Ely Porat 15:1–15:15

Approximating Longest Common Substring with k mismatches: Theory and Practice
Garance Gourdel, Tomasz Kociumaka, Jakub Radoszewski, and Tatiana Starikovskaya 16:1–16:15

String Factorizations Under Various Collision Constraints
Niels Grüttemeier, Christian Komusiewicz, Nils Morawietz, and Frank Sommer .. 17:1–17:14

k-Approximate Quasiperiodicity under Hamming and Edit Distance
Aleksander Kędzierski and Jakub Radoszewski 18:1–18:15

Longest Common Subsequence on Weighted Sequences
Evangelos Kipouridis and Kostas Tsichlas 19:1–19:15

Parameterized Algorithms for Matrix Completion with Radius Constraints
Tomohiro Koana, Vincent Froese, and Rolf Niedermeier 20:1–20:14

In-Place Bijective Burrows-Wheeler Transforms
Dominik Köppl, Daiki Hashimoto, Diptarama Hendrian, and Ayumi Shinohara ... 21:1–21:15

Genomic Problems Involving Copy Number Profiles: Complexity and Algorithms
Manuel Lafond, Binhai Zhu, and Peng Zou 22:1–22:15

Compressed Orthogonal Search on Suffix Arrays with Applications to Range LCP
Kotaro Matsuda, Kunihiko Sadakane, Tatiana Starikovskaya, and Masakazu Tateshita 23:1–23:13

Text Indexing and Searching in Sublinear Time
J. Ian Munro, Gonzalo Navarro, and Yakov Nekrich 24:1–24:15

Chaining with Overlaps Revisited
Veli Mäkinen and Kristoffer Sahlin 25:1–25:12

DAWGs for Parameterized Matching: Online Construction and Related Indexing Structures
Katsuhito Nakashima, Noriki Fujisato, Diptarama Hendrian, Yuto Nakashima, Ryo Yoshinaka, Shunsuke Inenaga, Hideo Bannai, Ayumi Shinohara, and Masayuki Takeda 26:1–26:14

On Extensions of Maximal Repeats in Compressed Strings
Julian Pape-Lange 27:1–27:13

Faster Binary Mean Computation Under Dynamic Time Warping
Nathan Schaar, Vincent Froese, and Rolf Niedermeier 28:1–28:13

Approximating Text-To-Pattern Distance via Dimensionality Reduction
Przemysław Uznański 29:1–29:11

Preface

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 31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020) held on June 17-19. The conference was planned to be held in Copenhagen, but due to the Covid-19 pandemic the conference was instead held online using Zoom. The conference program includes 28 contributed papers and three invited talks by Barna Saha (University of California Berkeley), Karl Bringmann (Max-Planck-Institut für Informatik), and Thore Husfeldt (IT University of Copenhagen and Lund University). For the second time, CPM includes the "Highlights of CPM" special session, for presenting the highlights of recent developments in combinatorial pattern matching. In this second edition we have invited Shay Golan to present his SODA 2020 paper "Locally consistent parsing for text indexing in small space", Tomasz Kociumaka to present his STOC 2019 paper "String synchronizing sets: sublinear-time BWT construction and optimal LCE data structure", and Paweł Gawrychowski (University of Wrocław, Poland) to present his paper "Computing quartet distance is equivalent to counting 4-Cycles".

The contributed papers were selected out of 49 submissions, corresponding to an acceptance ratio of about 57%. Each submission received at least three reviews. We thank the members of the Program Committee and all the additional external subreviewers that are listed below for their hard, invaluable, and collaborative effort that resulted in an excellent scientific program.

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, Qingdao, and Pisa. 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), 105 (CPM 2018), and 128 (CPM 2019). The entire submission and review process was carried out using the EasyChair conference system. We thank the CPM Steering Committee for their support and advice in this year's unusual circumstances.

31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020).
Editors: Inge Li Gørtz and Oren Weimann

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Programme Committee

Inge Li Gørtz (co-chair)
Technical University of , Denmark

Oren Weimann (co-chair)
University of Haifa, Israel

Amir Abboud
IBM, USA

Amihood Amir
Bar-Ilan University and Johns Hopkins University, Israel

Hideo Bannai
Kyushu University, Japan

Philip Bille
Technical University of Denmark, Denmark

Panagiotis Charalampopoulos
King's College London, UK

Debarati Das
University of Copenhagen, Denmark

Rolf Fagerberg
University of Southern Denmark, Denmark

Martin Farach-Colton
Rutgers University, USA

Gabriele Fici
Università di Palermo, Italy

Johannes Fischer
TU Dortmund, Germany

Elazar Goldenberg
The Academic College Of Tel Aviv-Yaffo, Israel

Roberto Grossi
Università di Pisa, Italy

Danny Hermelin
Ben Gurion University of the Negev, Israel

Artur Jez
University of Wroclaw, Poland

Dominik Kempa
University of California Berkeley, USA

Juha Kärkkäinen
University of Helsinki, Finland

Dominik Köppl
Kyushu University / JSPS, Japan

Avivit Levy
Shenkar College, Israel

Zsuzsanna Lipták
University of Verona, Italy

Giovanni Manzini
University of Eastern Piedmont, Italy

Shay Mozes
Interdisciplinary Center (IDC) Herzliya, Israel

Kunsoo Park
Seoul National University, South Korea

Solon Pissis
Centrum Wiskunde & Informatica (CWI), Netherlands

Nicola Prezza
University of Pisa, Italy

Jakub Radoszewski
University of Warsaw, Poland

Wojciech Rytter
University of Warsaw, Poland

Sharma Thankachan
University of Central Florida, USA

Przemyslaw Uznanski
University of Wrocław, Poland

Tomasz Waleń
University of Warsaw, Poland

31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020).
Editors: Inge Li Gørtz and Oren Weimann

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Reviewers

Aditi Dudeja

Aleksander Łukasiewicz

Alexandru Popa

Anish Mukherjee

Arnab Ganguly

Arseny Shur

Arturs Backurs

Bojian Xu

Brian Brubach

Daniel Gibney

Dekel Tsur

Diptarka Chakraborty

Eitan Kondratovsky

Evangelos Kipouridis

Garance Gourdel

Giovanna Rosone

Giovanni Pighizzini

Giulia Bernardini

Golnaz Badkobeh

Itai Boneh

Jonas Ellert

Julian Pape-Lange

Juliusz Straszyński

Kazuhiro Kurita

Manuel Lafond

Manuel Sorge

Marinella Sciortino

Mateusz Pawlik

Meirav Zehavi

Michał Gańczorz

Michelle Sweering

Pawel Gawrychowski

Ray Li

Rayan Chikhi

Ritu Kundu

Shay Golan

Shuhei Denzumi

Shunsuke Inenaga

Stéphane Vialette

Szymon Grabowski

Thierry Lecroq

Tomasz Kociumaka

Tomohiro I

Travis Gagie

Wiktor Zuba

Wing-Kai Hon

Yakov Nekrich

Yuto Nakashima

31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020).
Editors: Inge Li Gørtz and Oren Weimann
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

List of Authors

Aleksander Kedzierski
Alessio Conte
Amihood Amir
Arnab Ganguly
Ayelet Butman
Ayumi Shinohara
Binhai Zhu
Christian Komusiewicz
Daiki Hashimoto
Daniel Gibney
David Liben-Nowell
Dina Sokol
Diptarama Hendrian
Djamal Belazzougui
Dominik Köppl
Ely Porat
Evangelos Kipouridis
Frank Sommer
Gad M. Landau
Garance Gourdel
Giulia Bernardini
Gonzalo Navarro
Gregory Kucherov
Grigorios Loukides
Hideo Bannai
Huiping Chen
Ian Munro
Jakub Radoszewski
Julian Pape-Lange
Juliusz Straszyński
Katsuhito Nakashima
Kiran Tomlinson
Kostas Tsichlas
Kristoffer Sahlin
Kunihiko Sadakane
Leen Stougie
M. Oguzhan Kulekci
Manal Mohamed
Manuel Lafond
Masayuki Takeda
Matan Kraus
Michelle Sweering
Mitsuru Funakoshi
Nadia Pisanti
Niels Grüttemeier
Nils Morawietz
Noriki Fujisato
Panagiotis Charalampopoulos
Paola Bonizzoni
Patty Commins
Pawel Gawrychowski
Peng Zou
Przemysław Uznański
Rolf Niedermeier
Ryo Yoshinaka
Sahar Hooshmand
Samah Ghazawi
Sharma V. Thankachan
Shay Golan
Shay Mozes
Shoshana Marcus
Shunsuke Inenaga

31st Annual Symposium on Combinatorial Pattern Matching (CPM 2020).
Editors: Inge Li Gørtz and Oren Weimann
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Solon Pissis

Stav Ben Nun

Tatiana Starikovskaya

Tina Liu

Tomasz Kociumaka

Tomasz Waleń

Tomohiro Koana

Tsvi Kopelowitz

Veli Mäkinen

Vincent Froese

Wiktor Zuba

Wojciech Rytter

Yakov Nekrich

Yuto Nakashima