

32nd Annual Symposium on Combinatorial Pattern Matching

CPM 2021, July 5–7, 2021, Wrocław, Poland

Edited by

Paweł Gawrychowski

Tatiana Starikovskaya



Editors

Paweł Gawrychowski 

University of Wrocław, Poland
gawry@cs.uni.wroc.pl

Tatiana Starikovskaya

École normale supérieure, France
tat.starikovskaya@gmail.com

ACM Classification 2012

Theory of computation → Pattern matching

ISBN 978-3-95977-186-3

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/978-3-95977-186-3>.

Publication date

July, 2021

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 4.0 International license (CC-BY 4.0):
<https://creativecommons.org/licenses/by/4.0/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.2021.0

ISBN 978-3-95977-186-3

ISSN 1868-8969

<https://www.dagstuhl.de/lipics>

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*, Reykjavik University, IS and Gran Sasso Science Institute, IT)
- Christel Baier (TU Dresden, DE)
- Mikolaj Bojanczyk (University of Warsaw, PL)
- Roberto Di Cosmo (Inria and Université de Paris, FR)
- Faith Ellen (University of Toronto, CA)
- Javier Esparza (TU München, DE)
- Daniel Král' (Masaryk University - Brno, CZ)
- Meena Mahajan (Institute of Mathematical Sciences, Chennai, IN)
- Anca Muscholl (University of Bordeaux, FR)
- Chih-Hao Luke Ong (University of Oxford, GB)
- Phillip Rogaway (University of California, Davis, US)
- Eva Rotenberg (Technical University of Denmark, Lyngby, DK)
- Raimund Seidel (Universität des Saarlandes, Saarbrücken, DE and Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Wadern, DE)

ISSN 1868-8969

<https://www.dagstuhl.de/lipics>

■ Contents

Preface	
<i>Paweł Gawrychowski and Tatiana Starikovskaya</i>	0:vii
Programme Committee	
.....	0:ix
External Subreviewers	
.....	0:xi
List of Authors	
.....	0:xiii–xiv

Invited Talks

Repetitions in Strings: A “Constant” Problem	
<i>Hideo Bannai</i>	1:1–1:1
Computing Edit Distance	
<i>Michal Koucký</i>	2:1–2:1
On-Line Pattern Matching on D-Texts	
<i>Nadia Pisanti</i>	3:1–3:2

Regular Papers

Ranking Bracelets in Polynomial Time	
<i>Duncan Adamson, Vladimir V. Gusev, Igor Potapov, and Argyrios Deligkas</i>	4:1–4:17
The k -Mappability Problem Revisited	
<i>Amihod Amir, Itai Boneh, and Eitan Konradovsky</i>	5:1–5:20
Internal Shortest Absent Word Queries	
<i>Golnaz Badkobeh, Panagiotis Charalampopoulos, and Solon P. Pissis</i>	6:1–6:18
Constructing the Bijection and the Extended Burrows–Wheeler Transform in Linear Time	
<i>Hideo Bannai, Juha Kärkkäinen, Dominik Köppl, and Marcin Piątkowski</i>	7:1–7:16
Weighted Ancestors in Suffix Trees Revisited	
<i>Djamal Belazzougui, Dmitry Kosolobov, Simon J. Puglisi, and Rajeev Raman</i>	8:1–8:15
Constructing Strings Avoiding Forbidden Substrings	
<i>Giulia Bernardini, Alberto Marchetti-Spaccamela, Solon P. Pissis, Leen Stougie, and Michelle Sweering</i>	9:1–9:18
Gapped Indexing for Consecutive Occurrences	
<i>Philip Bille, Inge Li Gørtz, Max Rishøj Pedersen, and Teresa Anna Steiner</i>	10:1–10:19
Disorders and Permutations	
<i>Laurent Bulteau, Samuele Giraudo, and Stéphane Vialette</i>	11:1–11:15

32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021).

Editors: Paweł Gawrychowski and Tatiana Starikovskaya



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Computing Covers of 2D-Strings <i>Panagiotis Charalampopoulos, Jakub Radoszewski, Wojciech Rytter, Tomasz Waleń, and Wiktor Zuba</i>	12:1–12:20
A Fast and Small Subsampled R-Index <i>Dustin Cobas, Travis Gagie, and Gonzalo Navarro</i>	13:1–13:16
The Longest Run Subsequence Problem: Further Complexity Results <i>Riccardo Dondi and Florian Sikora</i>	14:1–14:15
Data Structures for Categorical Path Counting Queries <i>Meng He and Serikzhan Kazi</i>	15:1–15:17
Compressed Weighted de Bruijn Graphs <i>Giuseppe F. Italiano, Nicola Prezza, Blerina Sinimeri, and Rossano Venturini</i> ..	16:1–16:16
Optimal Construction of Hierarchical Overlap Graphs <i>Shahbaz Khan</i>	17:1–17:11
A Compact Index for Cartesian Tree Matching <i>Sung-Hwan Kim and Hwan-Gue Cho</i>	18:1–18:19
String Sanitization Under Edit Distance: Improved and Generalized <i>Takuya Mieno, Solon P. Pissis, Leen Stougie, and Michelle Sweering</i>	19:1–19:18
An Invertible Transform for Efficient String Matching in Labeled Digraphs <i>Abhinav Nellore, Austin Nguyen, and Reid F. Thompson</i>	20:1–20:14
R-enum: Enumeration of Characteristic Substrings in BWT-runs Bounded Space <i>Takaaki Nishimoto and Yasuo Tabei</i>	21:1–21:21
A Linear Time Algorithm for Constructing Hierarchical Overlap Graphs <i>Sangsoo Park, Sung Gwan Park, Bastien Cazaux, Kunsoo Park, and Eric Rivals</i>	22:1–22:9
Efficient Algorithms for Counting Gapped Palindromes <i>Andrei Popa and Alexandru Popa</i>	23:1–23:13
AWLCO: All-Window Length Co-Occurrence <i>Joshua Sobel, Noah Bertram, Chen Ding, Fatemeh Nargesian, and Daniel Gildea</i>	24:1–24:21
Optimal Completion and Comparison of Incomplete Phylogenetic Trees Under Robinson-Foulds Distance <i>Keegan Yao and Mukul S. Bansal</i>	25:1–25:23

■ Preface

The Annual Symposium on Combinatorial Pattern Matching (CPM) has by now over 30 years of tradition and is considered to be the leading conference for the community working on Stringology. The objective of the annual CPM meetings is to provide an international forum for research in combinatorial pattern matching and related applications such as computational biology, data compression and data mining, coding, information retrieval, natural language processing, and pattern recognition.

This volume contains the papers presented at the 32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021) held on July 5–7, 2021 in Wrocław, Poland (in a hybrid mode due to the continuing Covid-19 pandemic). The conference program includes 22 contributed papers and three invited talks by Hideo Bannai (M&D Data Science Center, Tokyo Medical and Dental University, Japan), Michal Koucký (Computer Science Institute of Charles University, Czech Republic), and Nadia Pisanti (University of Pisa, Italy and Erable Team INRIA, France). For the third time, CPM includes the “Highlights of CPM” special session, for presenting the highlights of recent developments in combinatorial pattern matching. In this third edition we invited Travis Gagie (CeBiB – Center for Biotechnology and Bioengineering, Chile and Dalhousie University, Canada) to present a J. of ACM 2020 paper by T. Gagie, G. Navarro, N. Prezza “Fully Functional Suffix Trees and Optimal Text Searching in BWT-Runs Bounded Space” and Panagiotis Charalampopoulos (The Interdisciplinary Center Herzliya, Israel) to present a FOCS 2020 paper by P. Charalampopoulos, T. Kociumaka, P. Wellnitz “Faster Approximate Pattern Matching: A Unified Approach”. The conference was preceded by a one-day student summer school taught by Jakub Radoszewski (University of Warsaw, Poland) and Martin Farach-Colton (Rutgers University, USA).

The contributed papers were selected out of 49 submissions, corresponding to an acceptance ratio of about 45%. Each submission received at least three reviews. We thank the members of the Program Committee and all the additional external subreviewers who 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, Pisa, and Copenhagen. 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), 128 (CPM 2019), and 161 (CPM 2020). 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.



■ Programme Committee

Golnaz Badkobeh
Goldsmiths University of London, UK

Frédérique Bassino
University Paris 13, France

Christina Boucher
University of Florida, USA

Laurent Bulteau
CNRS and Université Paris-Est
Marne-la-Vallée, France

Raphaël Clifford
University of Bristol, UK

Fabio Cunial
MPI-CBG, Germany

Funda Ergun
Indiana University, USA

Paweł Gawrychowski (co-chair)
University of Wrocław, Poland

Inge Li Gørtz
Technical University of Denmark

Stepan Holub
Charles University in Prague, Czech
Republic

Tomohiro I
Kyushu Institute of Technology, Japan

Shunsuke Inenaga
Kyushu University, Japan

Tomasz Kociumaka
University of California, Berkeley, USA

Christian Komusiewicz
Philipps-Universität Marburg, Germany

Dmitry Kosolobov
Ural Federal University, Russia

Gad M. Landau
University of Haifa, Israel

Florin Manea
University of Göttingen, Germany

Pierre Peterlongo
INRIA, France

Cinzia Pizzi
University of Padova, Italy

Leena Salmela
University of Helsinki, Finland

Srinivasa Rao Satti
Seoul National University, South Korea

Marinella Sciortino
University of Palermo, Italy

Braha-Riva Shalom
Shenkar College of Engineering and Design,
Israel

Tatiana Starikovskaya (co-chair)
Ecole normale supérieure, France

Yasuo Tabei
RIKEN, Japan

Tomasz Waleń
University of Warsaw, Poland

32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021).

Editors: Paweł Gawrychowski and Tatiana Starikovskaya



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

■ External Subreviewers

Hideo Bannai

Jérémy Barbay

Jacopo Borga

Bastien Cazaux

Davide Cenzato

Rayan Chikhi

Julien Clément

Nadia El-Mabrouk

Gabriele Fici

Arnab Ganguly

Samah Ghazawi

Niels Grüttemeier

Diptarama Hendrian

Jana Holznigenkemper

Varunkumar Jayapaul

Seungbum Jo

Dominik Köppl

Shunsuke Kanda

Dominik Kempa

Gunnar W. Klau

Shmuel Tomi Klein

Tore Koß

Christian Konrad

Maria Kosche

Thierry Lecroq

Inbok Lee

Avivit Levy

Noa Lewenstein

Antoine Limasset

Zsuzsanna Liptak

Bertrand Marchand

Nils Morawietz

Yuto Nakashima

Cyril Nicaud

Takaaki Nishimoto

Taku Onodera

Kunsoo Park

Karol Pokorski

Nicola Prezza

Simon Puglisi

Jakub Radoszewski

Gwenaël Richomme

Giuseppe Romana

Massimiliano Rossi

Paweł Rychlikowski

Arseny Shur

Stefan Siemer

Marcin Smulewicz

Dina Sokol

Frank Sommer

Teresa Anna Steiner

Juliusz Straszyński

Xiaorui Sun

Michelle Sweering

Nathan Wallheimer

Mathias Weller

Samson Zhou


Wiktor Zuba

■ List of Authors

Duncan Adamson
Amihood Amir
Golnaz Badkobeh
Hideo Bannai
Mukul S. Bansal
Djamal Belazzougui
Giulia Bernardini
Noah Bertram
Philip Bille
Itai Boneh
Laurent Bulteau
Bastien Cazaux
Panagiotis Charalampopoulos
Hwan-Gue Cho
Dustin Cobas
Argyrios Deligkas
Chen Ding
Riccardo Dondi
Travis Gagie
Daniel Gildea
Samuele Giraudo
Inge Li Gørtz
Vladimir Gusev
Meng He
Giuseppe F. Italiano
Juha Kärkkäinen
Serikzhan Kazi
Shahbaz Khan
Sung-Hwan Kim
Eitan Konradovskiy
Dominik Köppl
Dmitry Kosolobov
Alberto Marchetti-Spaccamela
Takuya Mieno
Fatemeh Nargesian
Gonzalo Navarro
Abhinav Nellore
Austin Nguyen
Takaaki Nishimoto
Sangsoo Park
Sung Gan Park
Kunsoo Park
Max Pedersen
Marcin Piątkowski
Solon Pissis
Alexandru Popa
Andrei Popa
Igor Potapov
Nicola Prezza
Simon Puglisi
Jakub Radoszewski
Rajeev Raman
Eric Rivals
Wojciech Rytter
Florian Sikora
Blerina Sinimeri
Joshua Sobel
Teresa Anna Steiner
Leen Stougie
Michelle Sweering
Yasuo Tabei
Reid Thompson

32nd Annual Symposium on Combinatorial Pattern Matching (CPM 2021).

Editors: Paweł Gawrychowski and Tatiana Starikovskaya

 Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

0:xiv **Authors**

Rossano Venturini

Stéphane Vialette

Tomasz Walen

Keegan Yao

Wiktor Zuba