europass

Carlo Tosoni

Nationality: Italian **Date of birth:** 16/04/1999 **Place of birth:** Assisi, Italy

Gender: Male **Phone number:** (+39) 3459913191

Email address: carlo.tosoni@unive.it

LinkedIn: https://www.linkedin.com/in/carlo-tosoni-42203b273/

Website: https://CarloTosoni99.github.io

Home: Via Baldassarre Longhena 32, 30175 Marghera (Italy)

ABOUT ME

Master's degree in computer science

EDUCATION AND TRAINING

PhD Student in Computer Science

Ca' Foscari University [08/09/2023 – Current]

Address: Via Torino 155, 30170 Venice (Italy) | **Level in EQF:** EQF level 8

Sorting is one of the most powerful techniques for enabling search on a particular data structure. For instance, using binary search is possible to retrieve an item from a sorted list in logarithmic time w.r.t. the list dimension. Similar techniques can be applied also to more convoluted data structures, like for instance finite-state automata. In fact, sorting the states of an automaton based on the strings reaching them enables the development of efficient indexes on the regular language recognized by the automaton itself. This indexing strategy naturally extends the renowned Burrows-Wheeler transform, originally devised for strings, to arbitrary finite-state automata.

During the PhD programme, we aim to significantly improve this indexing strategy by providing efficient algorithms and data structures to sort the states of automaton and to provide time-space efficient techniques to locate patterns with the resulting index. As the current state-of-the-art algorithm to compute such indexes has quadratic time complexity (w.r.t. the number of transitions of the input automaton) this indexing technique is still unfeasible in the realm of the so-called big data. Therefore, to development of a near-linear time algorithm represents a crucial step to make this indexing strategy a viable option for prominent research fields like bioinformatics, where the data dimension has soared exponentially over the past years.

The PhD programme is funded by the following grant of the European Union and it is supervised by professors Nicola Prezza and Ruben Becker of the Ca' Foscari University of Venice.

MSc in Computer Science, curriculum Big Data Technologies

University of Pisa [09/2021 – 07/2023]

Address: Largo Bruno Pontecorvo 3, 56127 Pisa (Italy) | **Final grade:** 110/110 cum laude |
Level in EQF: EQF level 7 | **Thesis:** Compressing the Burrows-Wheeler transform of finite-state automata using run-length encoding

The master's degree in computer science, curriculum Big Data Technologies, is offered by the Department of Computer Science at the University of Pisa. All the courses are provided entirely in English and are focused on the following topics.

• Design, analysis and implementation of advanced algorithms and data structures to efficiently solve combinatorial problems. Analysis of techniques and algorithms for implementing advanced databases. State-of-the-art techniques and paradigms for the analysis of genome sequences in bioinformatics.
• Principles and paradigms of machine learning. Analysis and implementation of algorithms for data mining.
• Megadata analysis, search engines, and information retrieval.
• High performance computing and parallel computing.

Achieved seven times the grade 30/30 cum laude. Weighted average at the end of studies 30.47/32.

BSc in Computer and Electonic Engineering

University of Perugia [09/2018 – 09/2021]

Address: Via Goffredo Duranti 93, 06125 Perugia (Italy) | **Final grade:** 110/110 con lode | **Level in EQF:** EQF level 6 | **Thesis:** Development of a decentralized blog using Solid technology

Courses attended
• Feedback control systems, Signal theory, Principles of automatic control, Internet basics.
• Algorithms and data structures, Database management systems, Programming (Java and C).
• Logic design and microcontrollers, Circuit theory, Electronic devices and technologies.

Secondary-School Degree

Liceo Scientifico Statale Annesso al Convitto Nazionale "Principe di Napoli" [09/2013 – 07/2018]

Address: Piazza Giacomo Matteotti 67, 06081 Assisi (Italy) | **Final grade:** 100/100

PUBLICATIONS

[2025]

Encoding Co-Lex Orders of Finite-State Automata in Linear Space

The Burrows-Wheeler transform (BWT) is a string transformation that enhances string indexing and compressibility. Cotumaccio and Prezza [SODA '21] extended this transformation to nondeterministic finite automata (NFAs) through co-lexicographic partial orders, i.e., by sorting the states of an NFA according to the co-lexicographic order of the strings reaching them. As the BWT of an NFA shares many properties with its original string variant, the transformation can be used to implement indices for locating specific patterns on the NFA itself. The efficiency of the resulting index is influenced by the width of the partial order on the states: the smaller the width, the faster the index. The most efficient index for arbitrary NFAs currently known in the literature is based on the coarsest forward-stable co-lex (CFS) order of Becker et al. [SPIRE '24]. In this paper, we prove that this CFS order can be encoded within linear space in the number of states in the automaton. The importance of this result stems from the fact that encoding such an order in linear space represents a big first step in the direction of building the index based on this order in near-linear time -- the biggest open research question in this context. The currently most efficient known algorithm for this task run in quadratic time in the number of transitions in the NFA and are thus infeasible to be run on very large graphs (e.g., pangenome graphs). At this point, a near-linear time algorithm is solely known for the simpler case of deterministic automata [Becker et al., ESA '23] and, in fact, this algorithmic result was enabled by a linear space encoding for deterministic automata [Kim et al., CPM '23].

Ruben Becker, Nicola Cotumaccio, Sung-Hwan Kim, Nicola Prezza, and Carlo Tosoni, submitted to CPM 2025

[2024]

Indexing Finite-State Automata Using Forward-Stable Partitions

An index on a finite-state automaton is a data structure able to locate specific patterns on the automaton's paths and consequently on the regular language accepted by the automaton itself. Cotumaccio and Prezza [SODA '21], introduced a data structure able to solve pattern matching queries on automata, generalizing the famous FM-index for strings of Ferragina and Manzini [FOCS '00]. The efficiency of their index depends on the width of a particular partial order of the automaton's states, the smaller the width of the partial order, the faster is the index. However, computing the partial order of minimal width is NP-hard. This problem was mitigated by Cotumaccio [DCC '22], who relaxed the conditions on the partial order, allowing it to be a partial preorder. This relaxation yields the existence of a unique partial preorder of minimal width that can be computed in polynomial time. In the paper at hand, we present a new class of partial preorders and show that they have the following useful properties: (i) they can be computed in polynomial time, (ii) their width is never larger than the width of Cotumaccio's preorders, and (iii) there exist infinite classes of automata on which the width of Cotumaccio's preorder is linearly larger than the width of our preorder.

Ruben Becker, Sung-Hwan Kim, Nicola Prezza, and Carlo Tosoni, SPIRE 2024

PROGRAMMING LANGUAGES

Programming languages and computer skills

Excellent knowledge of programming languages like Java, C++, C#, C. Worked also with HTML, Javascript, and CSS and frameworks such as React, Qwik, and Node.js to develop web applications. Good knowledge also with Python 3 and libraries like Pandas, PyTorch, and Keras to train/test machine learning models and to apply data mining techniques. Familiarity with the programs Unity and Blender for designing virtual realities and creating 3D models. Worked occasionally with Haskell, MATLAB, and Rust.

PEER REVIEWS FOR SCIENTIFIC PUBLICATIONS

Reviewer for journals

Reviewed articles for the following journals. Information Systems and Algorithms for Molecular Biology.

Sub-reviewer for conferences

Appointed sub-reviewer for the following conferences. Data Compression Conference (DCC), 2024 and 2025. Symposium on Combinatorial Pattern Matching (CPM), 2024. Conference on Wonderful Algorithms in Bioinformatics (WABI), 2024. Symposium on Simplicity in Algorithms (SOSA), 2024. International Workshop on Combinatorial Algorithms (IWOCA), 2025.

HONOURS AND AWARDS

[03/2022] British Council
IELTS Academic 7.0
Listening 7.5, Reading 7.5, Writing 6.5, Speaking 6.0.

[01/2021] Huawei Technologies co. ltd
Huawei HCIA Routing and Switching Certification
The HUAWEI HCIA Routing and Switching Certification certifies skills in the core technologies for Networking: IP network connectivity, TCP/IP technologies, Ethernet technologies such as STP and RSTP, VLAN and Link Aggregation and their implementation on Huawei switches.

TALKS

[05/02/2025]
Efficient Indexes for Pangenome Graphs through BWT-Based Data Structures

Workshop Data Structures in Bioinformatics (DSB) 2025.

The talk summarized the state-of-the-art techniques to extend the renowned Burrows-Wheeler transform from strings to arbitrary edge-labeled graphs. As this transform can be used to implement efficient graph indexes on the regular language recognized by the graph itself, it followed that this indexing strategy has a remarkable potential in the field of bioinformatics to index pangenome graphs.

TEACHING ACTIVITIES

[01/07/2024 – 05/07/2024]
Tutor for the summer schools SEAA (Scuola estiva di Algoritmi Avanzati)

Ca' Foscari University of Venice

ISTITUTIONAL ROLES

[12/2023 – 12/2025]

Representative of the PhD students in computer science

Appointed representative of the PhD students in computer science at the Ca' Foscari University of Venice for the academic years 2023/24 and 2024/25.

DRIVING LICENCE

Driving Licence: B

HOBBIES AND INTERESTS

Piano

I play piano since I was a child.

Autorizzo il trattamento dei miei dati personali presenti nel CV ai sensi dell'art. 13 d. lgs. 30 giugno 2003 n. 196 - "Codice in materia di protezione dei dati personali" e dell'art. 13 GDPR 679/16 - "Regolamento europeo sulla protezione dei dati personali".

Venezia, 11/03/2025

Carlo Tosoni