

Nicola Prezza: Scientific Curriculum

Personal data

Name	Nicola Prezza
Address	University of Pisa, Largo Bruno Pontecorvo, 3 - 56127 - Pisa (Italy)
Email	nicola.prezza@gmail.com
Webpage	https://nicolaprezza.github.io/
OrcID	http://orcid.org/0000-0003-3553-4953
H-index	6, according to Google Scholar.

Education

1/1/2014– 31/12/2016	Ph.D, Computer Science, Mathematics, and Physics, University of Udine. Ph.D. thesis: “Compressed Computation for Text Indexing”. Supervisor: Alberto Policriti.
26/9/2011– 18/10/2013	Master’s degree in Computer Science, University of Udine. Final grade: 110/110 cum laude. GPA: 3.90.
27/8/2012– 30/1/2013	Erasmus Exchange student, University of Southern Denmark, Odense.
25/9/2008– 14/12/2011	Bachelor’s degree in Computer Science, University of Udine. Final grade: 110/110 cum laude. GPA: 3.73.

Employments

1/3/2018–	PostDoc researcher, dept. of Computer Science, University of Pisa
1/1/2017– 28/3/2018	PostDoc researcher, AlgoLoG, DTU Compute, DTU, Denmark.
1/1/2014– 31/12/2016	Ph.D.-student, Computer Science, University of Udine.

Teaching Experience

15/9/2017– 8/12/2017	Ph.D course: <i>Compact data structures</i> . 5 ECTS. Technical University of Denmark, Kgs. Lyngby, Denmark. www2.compute.dtu.dk/courses/02951/
24/7/2017– 28/7/2017	<i>Aligning DNA sequences on compressed collections of genomes</i> (Summer school). The CODATA-RDA Research Data Science Applied workshops on Extreme sources of data, Bioinformatics and IoT/Big-Data Analytics, ICTP, Trieste. Teaching hours: 6.
Fall 2015	<i>Laboratorio di Architettura degli Elaboratori</i> . 3 ECTS. University of Udine.
Fall 2014	<i>Laboratorio di Architettura degli Elaboratori</i> . 3 ECTS. University of Udine.
23/10/2014	<i>Sequence Analysis for Epigenomics</i> . 2nd Bioinformatics Introductory Course, Polo d’Innovazione Genomica, Genetica e Biologia, Perugia. Teaching hours: 3.

PhD schools/visits abroad:

9/8/2016– 12/8/2016	Summer School on Bioinformatics Data Structures, Helsinki.
27/9/2015– 1/10/2015	School on Cancer, Evolution and Complexity, Como.
1/6/2016– 1/8/2016	Visiting PhD student, Courant institute, NYU, New York.
28/9/2014– 2/10/2014	School on Cancer, Systems and Complexity, Como.
24/3/2014– 28/3/2014	Visiting PhD student, SciLifeLab, Stockholm.

Awards:

20/06/2016	Best talk award, MatBio 2016, King’s college London.
------------	--

Program Committees

I am on the PC of the 29th Annual Symposium on Combinatorial Pattern Matching (CPM 2018) and I will chair the 13th Workshop on Compression, Text and Algorithms (WCTA 2018).

Reviewer activities

I reviewed more than 35 publications for international conferences and journals. Conferences include: CPM, SPIRE, LATA, DCC, SEA. Journals include: Algorithmica, Information Systems, Theoretical Computer Science, Journal of Discrete Algorithms, Algorithms for Molecular Biology.

Talks at international venues

Below I report a list of the main talks I have given at international conferences and workshops. I will moreover present our work *Dominik Kempa and Nicola Prezza*. “*At the Roots of Dictionary Compression: String Attractors*.” at the 50th Annual ACM Symposium on the Theory of Computing (STOC 2018).

Conferences

1. *In-Place Sparse Suffix Sorting*. Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), January 7-10, 2018, New Orleans.
2. *Optimal-time text indexing in BWT-runs bounded space*. Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), January 7-10, 2018, New Orleans.
3. *Succinct Partial Sums and Fenwick Trees*. String Processing and Information Retrieval (SPIRE), September 26-29, 2017, Palermo.
4. *From LZ77 to the Run-Length Burrows-Wheeler Transform, and Back*. Symposium on Combinatorial Pattern Matching (CPM), July 4-6, 2017, Warsaw.
5. *A Framework of Dynamic Data Structures for String Processing*. Symposium on Experimental Algorithms (SEA), June 21-23, 2017, King’s College London.
6. *Space-Efficient Re-Pair Compression*. Data Compression Conference (DCC), March 26-29, 2017, Snowbird (UT).
7. *A Randomized LCE Data Structure* (best student talk award). Mathematical foundations for bioinformatics (MatBio), July 20, 2016, King’s College London.
8. *Computing LZ77 in Run-Compressed Space*. Data Compression Conference (DCC), March 29 - April 1, 2016, Snowbird (UT).
9. *Fast Online Lempel-Ziv Factorization in Compressed Space*. String Processing and Information Retrieval (SPIRE), September 3, 2015, King’s college London.
10. *Average linear time and compressed space construction of the Burrows-Wheeler transform*. 9th International Conference on Language and Automata Theory and Applications (LATA), March 2-6, 2015, Nice, France.
11. *Hashing and Indexing: succinct data structures and smoothed analysis*. Symposium on Algorithms and Computation (ISAAC), December 15-17, 2014, Jeonju, Korea.

Workshops and Seminars

1. *String Attractors*. Workshop on Compression, Text and Algorithms (WCTA), September 29, 2017, Palermo, Italy.

2. *String Attractors*. NII Shonan Meeting: “Computation over Compressed Structured Data”, October 9-12, 2017, Shonan Village center, Japan.
3. *In-place Longest Common Extensions*. Dagstuhl seminar: Computation over Compressed Structured Data, October 28, 2016, Schloss Dagstuhl - Leibniz Center for Informatics.
4. *Indexing in repetition-aware space*. Dagstuhl seminar: Computation over Compressed Structured Data, October 26, 2016, Schloss Dagstuhl - Leibniz Center for Informatics.
5. *Algorithms for the compression of genomic big data*. BITS annual meeting, June 15-17, 2016, University of Salerno.
6. *Compressed Indexes for Populations of Genomes*. EPIGEN annual meeting, May 24-27 2016, Hotel Ergife, Rome.
7. *Space-efficient compression and indexing of genomic big data: theory and practice*. Workshop on Data Structures in Bioinformatics (DSB), February 23-24, 2016, University of Bielefeld.
8. *Aligning Nanopore Events on a Reference Genome*. Workshop on Compression, Text and Algorithms (WCTA), September 4, 2015, King’s college London.
9. *ONTRC: Bayesian Base-Calling of Nanopore Events for Portable Clinical Genomics*. IBM Thomas J. Watson Research Center, New York, July 30, 2015. Joint work with NYU’s Courant Institute of Mathematical Sciences (CIMS), and New York Genome Center (NYGC).
10. *Algorithms for the analysis of epigenetic data*. EPIGEN annual meeting, April 22, 2015, Hotel Ergife, Rome.
11. *Hashing and Indexing with succinct data structures*. Data Structures in Bioinformatics, December 8-9, 2014, Montpellier, France.
12. *Hashing and Indexing: succinct data structures and smoothed analysis*. December 2, 2014, University of Helsinki, Department of computer science.
13. *Analysis of capture-based bisulfite sequencing data with ERNE-BS5*. NETTAB, October 16, 2014, University of Torino, molecular Biotechnology Center (MBC).
14. *Differential methylation analysis of target enrichment protocols data with ERNE-BS5*. EPIGEN NGS and DATA ANALYSIS WORKSHOP, April 11, 2014, Istituto di Genomica Applicata, Udine.
15. *A tour to DNA methylation analysis via BS-seq reads alignment*. March 26, 2014, SciLifeLab, Stockholm.
16. *Fast randomized approximate string matching with succinct hash data structures*. BITS Annual Meeting, February 26-28, 2014, Rome.

Publications.

Publications on computer science conferences and journals (i.e. [PP17b] and all conferences except [PDFV⁺12]) follow the tradition of alphabetically sorted

author lists.

Peer reviewed journal publications.

- [PP17b] Alberto Policriti and Nicola Prezza. LZ77 Computation Based on the Run-Length Encoded BWT. *Algorithmica*, Jul 2017
- [PP15b] Alberto Policriti and Nicola Prezza. Fast randomized approximate string matching with succinct hash data structures. *BMC bioinformatics*, 16(9):S4, 2015
- [PVKP16] Nicola Prezza, Francesco Vezzi, Max Käller, and Alberto Policriti. Fast, accurate, and lightweight analysis of BS-treated reads with ERNE 2. *BMC bioinformatics*, 17(4):69, 2016
- [EWM⁺17] Karin Engström, Tomasz K Wojdacz, Francesco Marabita, Philip Ewels, Max Käller, Francesco Vezzi, Nicola Prezza, Joel Gruselius, Marie Vahter, and Karin Broberg. Transcriptomics and methylomics of CD4-positive T cells in arsenic-exposed women. *Archives of toxicology*, 91(5):2067–2078, 2017

Peer reviewed conference publications.

- [KP18] Dominik Kempa and Nicola Prezza. At the Roots of Dictionary Compression: String Attractors. In *Proceedings of the 50th Annual ACM Symposium on the Theory of Computing (STOC)*. To appear., 2018
- [Pre18] Nicola Prezza. In-Place Sparse Suffix Sorting. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, NA, USA, January 7-10, 2018*, pages 1496–1508. SIAM, 2018
- [GNP18b] Travis Gagie, Gonzalo Navarro, and Nicola Prezza. Optimal-Time Text Indexing in BWT-runs Bounded Space. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, NA, USA, January 7-10, 2018*. SIAM, 2018. Accepted for publication. Preprint available at: <https://arxiv.org/abs/1705.10382>

- [GNP18a] Travis Gagie, Gonzalo Navarro, and Nicola Prezza. On the approximation ratio of Lempel-Ziv parsing. In *Proc. 13th Latin American Symposium on Theoretical Informatics (LATIN)*, LNCS, 2018. To appear
- [BCPS17] Philip Bille, Anders Roy Christiansen, Nicola Prezza, and Fredrik Rye Skjoldjensen. Succinct Partial Sums and Fenwick Trees. In *International Symposium on String Processing and Information Retrieval (proceedings to appear)*. Springer, 2017
- [PP17a] Alberto Policriti and Nicola Prezza. From LZ77 to the Run-Length Encoded Burrows-Wheeler Transform, and Back. In *Annual Symposium on Combinatorial Pattern Matching (proceedings to appear)*. Springer, 2017
- [BGP17] Philip Bille, Inge Li Gørtz, and Nicola Prezza. Space-Efficient Re-Pair Compression. In *Data Compression Conference (DCC), 2017*, pages 171–180. IEEE, 2017
- [BCG⁺17] Djamel Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, and Mathieu Raffinot. Flexible Indexing of Repetitive Collections. In *Conference on Computability in Europe*, pages 162–174. Springer, 2017
- [Pre17] Nicola Prezza. A Framework of Dynamic Data Structures for String Processing. In *International Symposium on Experimental Algorithms (proceedings to appear)*. Springer, 2017
- [PP16] Alberto Policriti and Nicola Prezza. Computing LZ77 in run-compressed space. In *Data Compression Conference (DCC), 2016*, pages 23–32. IEEE, 2016
- [BCG⁺15] Djamel Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, and Mathieu Raffinot. Composite repetition-aware data structures. In *Annual Symposium on Combinatorial Pattern Matching*, pages 26–39. Springer, 2015
- [PP15a] Alberto Policriti and Nicola Prezza. Fast online Lempel-Ziv factorization in compressed space. In *International Symposium on String Processing and Information Retrieval*, pages 13–20. Springer, 2015

- [GPP15] Nicola Gigante, Alberto Policriti, and Nicola Prezza. Average linear time and compressed space construction of the Burrows-Wheeler transform. In *International Conference on Language and Automata Theory and Applications*, pages 587–598. Springer, 2015
- [PP14] Alberto Policriti and Nicola Prezza. Hashing and indexing: Succinct datastructures and smoothed analysis. In *International Symposium on Algorithms and Computation*, pages 157–168. Springer, 2014
- [PDFV⁺12] Nicola Prezza, Cristian Del Fabbro, Francesco Vezzi, Emanuele De Paoli, and Alberto Policriti. ERNE-BS5: aligning BS-treated sequences by multiple hits on a 5-letters alphabet. In *Proceedings of the ACM conference on bioinformatics, computational biology and biomedicine*, pages 12–19. ACM, 2012