Nicola Prezza

Scientific CV

▬▬▬ Education

Jan 1, 2014 – Dec 31, 2016	**Ph.D, Computer Science, Mathematics, and Physics**, *Department of Computer Science, University of Udine, Via delle Scienze, 206, 33100 Udine, Italy*, Thesis: *Compressed Computation for Text Indexing*, Final grade: passed cum laude.
Sep 26, 2011 – Oct 18, 2013	**Master's degree in Computer Science**, *Department of Computer Science, University of Udine, Via delle Scienze, 206, 33100 Udine, Italy*, Final grade: 110/110 cum laude. GPA: 3.90.
Aug 27, 2012 – Jan 30, 2013	**Erasmus Exchange student**, *University of Southern Denmark, Odense, Denmark*, Subjects: bioinformatics, combinatorial optimization, statistical simulation.
Sep 25, 2008 – Dec 14, 2011	**Bachelor's degree in Computer Science**, *Department of Computer Science, University of Udine, Via delle Scienze, 206, 33100 Udine, Italy*, Final grade: 110/110 cum laude. GPA: 3.73.

▬▬▬ Employments

Mar 1, 2018 –	**PostDoc researcher**, *Department of Computer Science, University of Pisa, Italy*, Project: *application of compressed indexes to aligment-free methods in bioinformatics*.
jan 1, 2017 – Mar 28, 2018	**PostDoc researcher**, *AlgoLoG, DTU Compute, Technical University of Denmark, Denmark*, Project: *Algorithms for the compression of big repetitive data*.
Jan 1, 2014 – Dec 31, 2016	**Ph.D.-student in Computer Science**, *University of Udine, Italy*, Project: *Compressed Computation for Text Indexing*.

▬▬▬ Teaching Experience

Sep 15, 2017 – Dec 8, 2017	**Compact data structures (Ph.D course)**, *Technical University of Denmark, Kgs. Lyngby, Denmark.* `www2.compute.dtu.dk/courses/02951`, 5 ECTS.
Jul 24, 2017 – Jul 28, 2017	**Aligning DNA sequences on compressed collections of genomes (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	**Laboratorio di Architettura degli Elaboratori**, *University of Udine*, 3 ECTS.
Fall 2014	**Laboratorio di Architettura degli Elaboratori**, *University of Udine*, 3 ECTS.
Oct 23, 2014	**Sequence Analysis for Epigenomics**, *2nd Bioinformatics Introductory Course, Polo d'Innovazione Genomica, Genetica e Biologia, Perugia. Teaching hours: 3.*

Visits abroad

Jun 1, 2016 – Aug 1, 2016	**Visiting PhD student**, *Courant institute, NYU, New York*, Project: *Oxford Nanopore base calling via Hidden Markov Models and compressed text indexes*.
mar 24, 2014 – Mar 28, 2014	**Visiting PhD student**, *SciLifeLab, Stockholm*, Project: *Development of a tool for DNA methylation analysis*.

Awards

Jan 1, 2014	**PhD fellowship**, *Department of Computer Science, University of Udine, Italy*.
Jun 20, 2016	**Best talk award**, *MatBio 2016, King's college London*, Talk title: *A Randomized In-Place LCE Data Structure*.

Academic Activities

Program Committees

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

Review activities

I reviewed more than 40 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 Conferences

In addition to the talks listed below, I will present our paper *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).

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.

Talks at 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 Mathematcal 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

In addition to the works listed below, I submitted seven additional papers (to top-rated journals and conferences) that are waiting for peer-review. For a complete list including submitted papers, please refer to `www.researchgate.net/profile/Nicola_Prezza`.

2018

International Conferences

1. Dominik Kempa and Nicola Prezza. **At the Roots of Dictionary Compression: String Attractors**. To appear in *proceedings of the 50th Annual ACM Symposium on the Theory of Computing (STOC)*, June 25-29, 2018 in Los Angeles, CA.
2. Travis Gagie, Gonzalo Navarro, and Nicola Prezza. **Optimal-Time Text Indexing in BWT-runs Bounded Space**. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Pages 1459-1477.
3. Nicola Prezza. **In-Place Sparse Suffix Sorting**. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Pages 1496-1508.
4. Travis Gagie, Gonzalo Navarro, and Nicola Prezza. **On the Approximation Ratio of Lempel-Ziv Parsing**. To appear in *Proceedings of LATIN 2018*.
5. Nicola Prezza, Nadia Pisanti, Marinella Sciortino, and Giovanna Rosone. **Detecting Mutations by eBWT**. To appear in *Proceedings of WABI 2018*.

2017

PhD thesis

1. Nicola Prezza. **Compressed Computation for Text Indexing**. *PhD thesis. University of Udine, 2017*.

International Journals

1. 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, no. 5 (2017): 2067-2078*.
2. Alberto Policriti and Nicola Prezza. **LZ77 computation based on the run-length encoded BWT**. *Algorithmica* (2017): 1-26.

International Conferences

1. Philip Bille, Inge Li Gørtz, and Nicola Prezza. **Space-Efficient Re-Pair Compression**. Proceedings of *Data Compression Conference (DCC)*, 2017. IEEE, 2017.
2. Nicola Prezza. **A Framework of Dynamic Data Structures for String Processing**. Proceedings of the *16th International Symposium on Experimental Algorithms (SEA)*, 2017.

3. Djamal Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, and Mathieu Raffinot. **Flexible indexing of repetitive collections**. In *Conference on Computability in Europe (CiE)*, pp. 162-174. Springer, Cham, 2017.

4. Philip Bille, Anders Roy Christiansen, Nicola Prezza, and Frederik Rye Skjoldjensen. **Succinct Partial Sums and Fenwick Trees**. In *International Symposium on String Processing and Information Retrieval (SPIRE)*, pp. 91-96. Springer, Cham, 2017.

5. Alberto Policriti and Nicola Prezza. **From LZ77 to the Run-Length Encoded Burrows-Wheeler Transform, and Back**. In *proceedings of the 28th Annual Symposium on Combinatorial Pattern Matching (CPM)*, 2017.

2016

International Journals

1. Nicola Prezza, Francesco Vezzi, Max Käller, and Alberto Policriti. **Fast, accurate, and lightweight analysis of BS-treated reads with ERNE 2**. *BMC bioinformatics* 17, no. 4 (2016): 69.

International Conferences

1. Alberto Policriti and Nicola Prezza. **Computing LZ77 in run-compressed space**. In *Data Compression Conference (DCC)*, 2016, pp. 23-32. IEEE, 2016.

2015

International Journals

1. Alberto Policriti and Nicola Prezza. **Fast randomized approximate string matching with succinct hash data structures**. *BMC bioinformatics* 16, no. 9 (2015): S4.

International Conferences

1. Djamal Belazzougui, Fabio Cunial, Travis Gagie, Nicola Prezza, and Mathieu Raffinot. **Composite repetition-aware data structures**. In *Annual Symposium on Combinatorial Pattern Matching (CPM)*, pp. 26-39. Springer, Cham, 2015.

2. Alberto Policriti and Nicola Prezza. **Fast online lempel-ziv factorization in compressed space**. In *International Symposium on String Processing and Information Retrieval (SPIRE)*, pp. 13-20. Springer, Cham, 2015.

3. Alberto Policriti, Nicola Gigante, 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 (LATA)*, pp. 587-598. Springer, Cham, 2015.

University of Pisa, Largo Bruno Pontecorvo 3 – 56127, Pisa – Italy
+39 3471530301 • nicola.prezza@gmail.com
nicolaprezza.github.io • it.linkedin.com/in/nicola-prezza-573898b2
github.com/nicolaprezza   *5/6*

2014

International Conferences

1. Alberto Policriti and Nicola Prezza. **Hashing and indexing: Succinct data structures and smoothed analysis**. In *International Symposium on Algorithms and Computation (ISAAC)*, pp. 157-168. Springer, Cham, 2014.

2012

International Conferences

1. Nicola Prezza, Cristian Del Fabbro, Francesco Vezzi, Emanuale 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 (ACM-BCB)*, pp. 12-19. ACM, 2012.

University of Pisa, Largo Bruno Pontecorvo 3 – 56127, Pisa – Italy
+39 3471530301 • nicola.prezza@gmail.com
nicolaprezza.github.io • it.linkedin.com/in/nicola-prezza-573898b2
github.com/nicolaprezza *6/6*