

**UNIVERSITÀ DEGLI STUDI DI MILANO**

selezione pubblica per n. 1 posto di Ricercatore a tempo determinato ai sensi dell'art.24, comma 3, lettera a) della Legge 240/2010 per il settore concorsuale 01/B1 - Informatica, settore scientifico-disciplinare INF/01 - Informatica presso il Dipartimento di INFORMATICA "GIOVANNI DEGLI ANTONI", (avviso bando pubblicato sulla G.U. n. G.U. 46 del 11/06/2021) Codice concorso 4759

## **Giulio Ermanno Pibiri**

### **CURRICULUM VITAE**

**PERSONAL INFORMATION**

<b>SURNAME</b>	PIBIRI
<b>NAME</b>	GIULIO ERMANNO
<b>DATE OF BIRTH</b>	13/07/1990

**EDUCATION**

- **01/11/2015 - 31/10/2018 — PhD in Computer Science (INF/01)**  
University of Pisa, Pisa, Italy  
Thesis: *Space- and Time-Efficient Data Structures for Massive Datasets* (Defended on 08/03/2019)  
Grade: Excellent  
Supervisor: Rossano Venturini (<https://rossanoventurini.github.io>)
- **2012 - 2014 — Master Degree in Computer Science and Networking (class LM18)**  
University of Pisa and Scuola Superiore Sant'Anna, Pisa, Italy  
Thesis: *Dynamic Elias-Fano Encoding* (Defended on 06/03/2015)  
Grade: 110/110 summa cum laude  
Supervisor: Rossano Venturini (<https://rossanoventurini.github.io>)
- **2009 - 2012 — Bachelor Degree in Computer Engineering (class L08)**  
University of Florence, Florence, Italy  
Thesis: *Quantum Computation and Grover's Algorithm* (Defended on 09/10/2012)  
Grade: 110/110 summa cum laude  
Supervisor: Gabriele Vezzosi (<http://www.dma.unifi.it/~vezzosi>)
- **2004 - 2009 — High School Diploma.**  
Liceo Scientifico Statale Guido Castelnuovo, Florence, Italy  
Grade: 100/100.

**RESEARCH INTERESTS****Keywords**

Data Structures, Data Compression, Indexing, Efficiency

**Short Description**

The research activity focuses on devising compressed data structures and algorithms to index and search large quantities of data. The proposed solutions are available as research papers and optimized software libraries (written in C++).

**Research Problems**

- succinct and dynamic ordered sets of integers (CPM 2017)
- inverted index compression (TOIS 2017, EBDT 2018, WSDM 2019, TKDE 2019, CSUR 2020)

- indexing and estimation of language models (SIGIR 2017, TOIS 2019)
- indexing of semantic relations (TKDE 2020, ICDE 2021)
- bitmap compression (DCC 2021)
- query auto-completion (SIGIR 2020)
- rank/select queries (INFOSYS 2021)
- prefix-sums (SPE 2020)
- minimal perfect hashing (SIGIR 2021)

## **RESEARCH ACTIVITY AND GRANTS**

- **15/03/2021 - present – Postdoctoral Research Fellow in Computer Science**  
**National Research Council of Italy (CNR), Pisa, Italy**  
 Institute of Science and Information Technologies “A. Faedo” (ISTI)

Research grant issued on the European project **ACCORDION** with theme “Tecniche algoritmiche per compressione, indicizzazione e ricerca di grandi quantità di dati e progettazione di relative librerie software open source” (Protocollo n. 0000901/2021, 09/03/2021, ISTI 004/2021 - PI).

Publications: SIGIR 2021

- **01/11/2018 - 28/02/2021 – Postdoctoral Research Fellow in Computer Science**  
**National Research Council of Italy (CNR), Pisa, Italy**  
 Institute of Science and Information Technologies “A. Faedo” (ISTI)

Research grant issued on the European project **BIGDATAGRAPES** with theme “Compressione, indicizzazione e ricerca su grandi collezioni di dati semantici” (Protocollo n. 0003847, 24/10/2018, ISTI 014/2018 - PI).

Publications: TKDE 2019, TKDE 2020, SIGIR 2020, SPE 2020, INFOSYS 2021, DCC 2021, ICDE 2021

- **01/11/2015 - 31/10/2018 – PhD Student in Computer Science**  
**University of Pisa, Pisa, Italy**  
 Supervisor: Rossano Venturini (<https://rossanoventurini.github.io>)  
 Thesis: *Space- and Time-Efficient Data Structures for Massive Datasets*  
 Worked on inverted indexing, compressed language models, and tries.  
 The thesis is based on the following publications.

- “On Optimally Partitioning Variable-Byte Codes” (TKDE 2019)
- “Handling Massive N-Gram Datasets Efficiently” (TOIS 2019)
- “Fast Dictionary-based Compression for Inverted Indexes” (WSDM 2019)
- “Inverted Index Compression” (EBDT 2018)
- “Efficient Data Structures for Massive N-Gram Datasets” (SIGIR 2017)
- “Dynamic Elias-Fano Representation” (CPM 2017)
- “Clustered Elias-Fano Indexes” (TOIS 2017)

Part of the research was conducted abroad:

- **01/05/2018 - 01/10/2018**

**The University of Melbourne, School of Computing and Information Systems, Melbourne, Australia**

Supervisor: Alistair Moffat (<https://people.eng.unimelb.edu.au/ammoffat>)

Worked on fast dictionary-based decoding of compressed inverted index data.

Reference publication: “Fast Dictionary-based Compression for Inverted Indexes” (WSDM 2019).

- **01/04/2018 - 30/04/2018**

**RIKEN Advanced Intelligence Project (AIP), Tokyo, Japan**

Supervisor: Yasuo Tabei (<https://sites.google.com/site/yasuotabei>)

Worked on various problems, such as, string similarity search, trie indexing, rank/select indexes, and sparse matrix multiplication.

Reference publication: “Rank/Select Queries over Mutable Bitmaps” (INFOSYS 2021).

## **PROJECTS**

- **15/03/2021 - present**  
**National Research Council of Italy (CNR), Pisa, Italy**  
Institute of Science and Information Technologies “A. Faedo” (ISTI)

European project **ACCORDION** with theme “Tecniche algoritmiche per compressione, indicizzazione e ricerca di grandi quantità di dati e progettazione di relative librerie software open source” (Protocollo n. 0000901/2021, 09/03/2021, ISTI 004/2021 - PI).

Role: Principal Investigator

- **01/11/2018 - 28/02/2021**  
**National Research Council of Italy (CNR), Pisa, Italy**  
Institute of Science and Information Technologies “A. Faedo” (ISTI)

European project **BIGDATAGRAPES** with theme “Compressione, indicizzazione e ricerca su grandi collezioni di dati semantici” (Protocollo n. 0003847, 24/10/2018, ISTI 014/2018 - PI).

Role: Principal Investigator

- **01/06/2017 - 31/10/2018**  
**National Research Council of Italy (CNR), Pisa, Italy**  
Institute of Science and Information Technologies “A. Faedo” (ISTI)

European project **Large-scale Indie Gaming Analytics (LIGA)**.

Role: Principal Developer

## **OTHER WORK EXPERIENCE**

- **01/04/2015 - 01/07/2015 – Software Engineer Intern**  
**IBM, Rome, Italy**  
Supervisor: Alessio Fioravanti

## **TRAINING ACTIVITY**

- **05/03/2016 - 11/03/2016 – BISS 2016 - Bertinoro International Spring School 2016**

I attended (and passed the exam) the 3 courses offered by the school (each consisting of 13 hours of lectures):

- **Advanced Topics in Programming Languages**  
Prof. Giuseppe Castagna - Université Paris Diderot - Paris 7
- **Models and Languages for Service-Oriented and Cloud Computing**  
Prof. Gianluigi Zavattaro - University of Bologna
- **Algorithmic methods for mining large graphs**  
Prof. Aristides Gionis, Aalto University

## **TEACHING ACTIVITY**

- **02/2020 - 06/2020 – Teacher for “Algorithmics and Laboratory” - Corso B, code 008AA, 3 CFU, Bachelor Degree in Computer Science, University of Pisa, Italy**
- **02/2019 - 06/2019 – Assistant for “Algorithmics and Laboratory” - Corso A, code 008AA, 3 CFU, Bachelor Degree in Computer Science, University of Pisa, Italy**

- 09/2018 - 12/2018 – Assistant for “Competitive Programming and Contests”, code 645AA, 6 hours, Master Degree in Computer Science, University of Pisa, Italy
- 09/2017 - 12/2017 – Assistant for “Competitive Programming and Contests”, code 645AA, 6 hours, Master Degree in Computer Science, University of Pisa, Italy
- 09/2016 - 12/2016 – Teacher for “Algorithmics and Laboratory” - Corso di recupero, code 008AA, 3 CFU, Bachelor Degree in Computer Science, University of Pisa, Italy
- 02/2016 - 06/2016 – Assistant for “Algorithmics and Laboratory “- Corso A, code 008AA, 3 CFU, Bachelor Degree in Computer Science, University of Pisa, Italy

## **ORGANIZING COMMITTEES**

- 2020 – Member of the Organizing Committee of the 28-th edition of the Annual European Symposium on Algorithms (ESA 2020).
- 2019 – Member of the Organizing Committee of the 30-th edition of the International Symposium on Combinatorial Pattern Matching (CPM 2019).
- 2017 – Member of the Organizing Committee of the 24-th International Symposium on String Processing and Information Retrieval (SPIRE 2017).
- 2016 – Student volunteer for the organization of the 39-th ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016).

## **PROGRAM COMMITTEES**

- 2021 – Member of the Program Committee of the 15th IEEE International Conference on Application of Information and Communication Technologies (AICT 2021).
- 2021 – Member of the Program Committee of the 30-th ACM International Conference on Information and Knowledge Management (CIKM 2021).
- 2021 – Member of the Program Committee of the 44-th ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021).
- 2021 – Member of the Program Committee of the 43-rd European Conference on Information Retrieval (ECIR 2021).
- 2020 – Member of the Program Committee of the 14-th International ACM Conference on Web Search and Data Mining (WSDM 2021).
- 2020 – Member of the Program Committee of the 29-th ACM International Conference on Information and Knowledge Management (CIKM 2020).
- 2020 – Member of the Program Committee of the 43-rd ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020).
- 2019 – Member of the Program Committee of the 42-nd ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).
- 2018 – Member of the Program Committee of the 2-nd Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis (KG4IR), in conjunction with ACM SIGIR 2018.

## **REVIEWING ACTIVITY**

As of 2016, I am or have been an anonymous reviewer for the following international conferences/journals.

### **Conferences**

- SIGIR – ACM Conference on Research and Development in Information Retrieval
- WSDM – ACM Conference on Web Search and Data Mining
- WWW – The Web Conference
- CIKM – ACM Conference on Information and Knowledge Management
- CPM – Annual Symposium on Combinatorial Pattern Matching
- DCC – IEEE Data Compression Conference
- ECIR – European Conference on Information Retrieval
- ESA – European Symposium on Algorithms

### **Journals**

- TALG – ACM Transactions on Algorithms
- INFOSYS – Information Systems
- SPE – Software: Practice and Experience
- SPIRE – String Processing and Information Retrieval
- Algorithmica

## **TALKS**

### **Conferences**

- 07/2021 – PTHash: Revisiting FCH Minimal Perfect Hashing. SIGIR conference presentation. Montreal, Canada (Virtual event).
- 04/2021 – Compressed Indexes for Fast Search of Semantic Data. ICDE conference presentation. Chania, Crete, Greece (Virtual event).
- 03/2021 – Fast and Compact Set Intersection through Recursive Universe Partitioning. DCC conference presentation. Snow Bird, USA (Virtual event).
- 27/07/2020 – Efficient and Effective Query Auto-Completion. SIGIR conference presentation. Xi'An, China (Virtual event).
- 17/09/2019 – Compressed Indexes for Fast Search of Semantic Data. IIR conference presentation. Department of Information Engineering, Padova, Italy.
- 12/02/2019 – Fast Dictionary-based Compression for Inverted Indexes. WSDM conference presentation. Melbourne, Australia.
- 10/08/2017 – Efficient Data Structures for Massive N-Gram Datasets. SIGIR conference presentation. Keio Plaza Hotel, Tokyo, Japan.
- 06/07/2017 – Dynamic Elias-Fano Representation. CPM conference presentation. University Library of Warsaw, Warsaw, Poland.
- 06/06/2017 – Efficient Data Structures for Massive N-Gram Datasets. IIR conference presentation. Università della Svizzera Italiana, Lugano, Switzerland.

### **Seminars**

- 04/03/2021 – Efficiency for Real-World Applications. Seminar. ISTI-CNR, Pisa, Italy (Virtual event).
- 07/06/2019 – Ordered Set Problems. Seminar. ISTI-CNR, Pisa, Italy.
- 08/03/2019 – Space- and Time-Efficient Data Structures. PhD thesis defense. Department of Computer Science, Pisa, Italy.

- 01/02/2019 — Indexing Compressed Data for Fast Retrieval. PhD Students' Event. Department of Computer Science, Pisa, Italy.
- 29/10/2018 — Effective Web Graph Representations. Seminar. Department of Computer Science, Pisa, Italy.
- 17/05/2018 — On Optimally Partitioning Variable-Byte Index Data. Seminar. RMIT University, Melbourne, Australia.
- 10/04/2018 — Elias-Fano Encoding: a powerful tool for data structure design. Seminar. RIKEN AIP, Tokyo, Japan.
- 21/06/2016 — Elias-Fano Encoding: succinct representation of monotone integer sequences with search operations. Seminar. Department of Computer Science, Pisa, Italy.

## Other

- 15/11/2018 — Space- and Time-Efficient Data Structures. PhD research results. Department of Computer Science, Pisa, Italy.
- 10/10/2017 — Space- and Time-Efficient Data Structures. PhD research results. Department of Computer Science, Pisa, Italy.
- 17/10/2016 — Space- and Time-Efficient Data Structures. PhD thesis proposal. Department of Computer Science, Pisa, Italy.

## AWARDS AND GRANTS

- 2020 — Young Researcher Award issued by ISTI-CNR.
- 2017 — SIGIR Student Travel Grant issued by ACM SIGIR.
- 2015 — PhD Scholarship issued by the University of Pisa, Department of Computer Science.
- 2015 — Master Degree Award: Best Performance a.y. 2013/2014 issued by Scuola Superiore Sant'Anna.
- 2015 — Best Master Thesis Award in Theoretical Computer Science, issued by the Italian chapter of the European Association for Theoretical Computer Science (EATCS).

## SOFTWARE

All software is open-source and available at <https://github.com/jermp>.

### Data Structures

Efficient C++ implementations of the following data structures (see also related publications):

- Inverted Indexes (TOIS 2017, TKDE 2019, WSDM 2019, SIGIR 2020, CSUR 2020)
- Tries (SIGIR 2017, TOIS 2019, TKDE 2020)
- Compressed Bitmaps (DCC 2021)
- Mutable Bitmaps with Rank/Select (INFOSYS 2021)
- Segment-Trees and Fenwick-Trees (SPE 2020)
- Minimal Perfect Hash Functions (SIGIR 2021)

Amore detailed list follows below.

- <https://github.com/jermp/pthash>  
PTHash: Fast and compact minimal perfect hash functions.  
Reference publications: SIGIR 2021, arXiv 2106.02350 2021.
- [https://github.com/jermp/mutable\\_rank\\_select](https://github.com/jermp/mutable_rank_select)

Mutable bitmaps with support for Rank and Select queries.  
Reference publication: INFOSYS 2021.

- <https://github.com/jermp/psds>

A range of tree-shaped data structures for maintaining prefix-sums, including:

- binary Segment-Tree (top-down and bottom-up),
- b-ary Segment-Tree,
- Fenwick-Tree,
- b-ary Fenwick-Tree,
- blocked Fenwick-Tree,
- truncated Fenwick-Tree.

Reference publication: SPE 2020.

- <https://github.com/jermp/autocomplete>

Efficient and effective autocompletion framework, based on forward/inverted indexes, succinct RMQ, and string dictionaries (Front-Coding and tries).

Reference publication: SIGIR 2020.

- [https://github.com/jermp/2i\\_bench](https://github.com/jermp/2i_bench)

A benchmarking suite for inverted index data structures, featuring the following compressors:

- Elias-Fano and partitioned Elias-Fano,
- Opt-PFor-Delta,
- Binary Interpolative,
- QMX,
- Simple family,
- Variable-Byte family, including Opt-VByte,
- Gamma, Delta, Rice, Zeta,
- DINT.

Reference publication: CSUR 2020.

- [https://github.com/jermp/interpolative\\_coding](https://github.com/jermp/interpolative_coding)

An efficient implementation of the Binary Interpolative Coding algorithm.

- [https://github.com/jermp/s\\_indexes](https://github.com/jermp/s_indexes)

Compressed bitmap indexes that support fast intersection and union.

Reference publication: DCC 2021.

- [https://github.com/jermp/rdf\\_indexes](https://github.com/jermp/rdf_indexes)

Trie-based indexes for semantic data like RDF triples.

Reference publication: TKDE 2020.

- <https://github.com/jermp/dint>

DINT: fast and compact dictionary-based decoder for inverted lists.

Reference publication: WSDM 2019.

- [https://github.com/jermp/opt\\_vbyte](https://github.com/jermp/opt_vbyte)

Optimal partitioning of inverted lists compressed using binary vectors and point-wise encoders, like Variable-Byte.

Reference publication: TKDE 2019.

- <https://github.com/jermp/tongrams>

Fast language model queries and estimation in compressed space.

Reference publications: SIGIR 2017, TOIS 2019.

- [https://github.com/jermp/clustered\\_elias\\_fano\\_indexes](https://github.com/jermp/clustered_elias_fano_indexes)

Clustered Elias-Fano inverted indexes.

Reference publication: TOIS 2017.

## Miscellanea

- <https://github.com/jermp/essentials>

A C++ library providing essential core utilities for data structure design and benchmarking.

More precisely:

- benchmarking facilities, including: messages displaying local time, configurable timer class, function to prevent code elision by compiler, simple creation and printing of json documents;
- functions to serialize-to and load-from disk data structures,
- functions to compute the number of bytes consumed by data structures,
- support for creating, removing, and iterate inside directories,
- transparent support for contiguous memory allocation.

- [https://github.com/jermp/cmd\\_line\\_parser](https://github.com/jermp/cmd_line_parser)

Command line parser for C++17. It offers all handy features in just 150 lines of code.

- [https://github.com/jermp/mm\\_file](https://github.com/jermp/mm_file)

A self-contained, header-only, implementation of memory-mapped files in C++ for both reading and writing.

## PUBLICATIONS

**Average number of co-authors per publication:  $20/18 = 1.111$**

(PhD Thesis excluded.)

### Journal Papers

- 2021 Giulio Ermanno Pibiri and Shunsuke Kanda, Rank/Select Queries over Mutable Bitmaps. Information Systems (INFOSYS), pages 15.  
Scimago Rating: **Q2**  
DOI: 10.1016/j.is.2021.101756
- 2020 Giulio Ermanno Pibiri and Rossano Venturini, Techniques for Inverted Index Compression. ACM Computing Surveys (CSUR), pages 36.  
Scimago Rating: **Q1**  
DOI: 10.1145/3415148
- 2020 Giulio Ermanno Pibiri and Rossano Venturini, Practical Trade-Offs for the Prefix-Sum Problem. Software: Practice and Experience (SPE), pages 29.  
Scimago Rating: **Q2**  
DOI: 10.1002/spe.2918
- 2020 Raffaele Perego, Giulio Ermanno Pibiri and Rossano Venturini, Compressed Indexes for Fast Search of Semantic Data. IEEE Transactions on Knowledge and Data Engineering (TKDE), pages 12.  
Scimago Rating: **Q1**  
DOI: 10.1109/TKDE.2020.2966609
- 2019 Giulio Ermanno Pibiri and Rossano Venturini, On Optimally Partitioning Variable-Byte Codes. IEEE Transactions on Knowledge and Data Engineering (TKDE), pages 12.  
Scimago Rating: **Q1**  
DOI: 10.1109/TKDE.2019.2911288
- 2019 Giulio Ermanno Pibiri and Rossano Venturini, Handling Massive N-Gram Datasets Efficiently. ACM Transactions on Information Systems (TOIS), pages 41.  
Scimago Rating: **Q1**  
DOI: 10.1145/3302913
- 2017 Giulio Ermanno Pibiri and Rossano Venturini, Clustered Elias-Fano Indexes. ACM Transactions on Information Systems (TOIS), volume 2, pages 33.  
Scimago Rating: **Q1**  
DOI: 10.1145/3052773



## Conference Papers

- 2021 Giulio Ermanno Pibiri and Roberto Trani, PTHash: Revisiting FCH Minimal Perfect Hashing. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.  
GGs Rating: A++  
DOI: To Appear.
- 2021 Giulio Ermanno Pibiri, Fast and Compact Set Intersection through Recursive Universe Partitioning. IEEE Data Compression Conference (DCC), pages 10.  
GGs Rating: A-  
DOI: 10.1109/DCC50243.2021.00037
- 2020 Simon Gog, Giulio Ermanno Pibiri and Rossano Venturini, Efficient and Effective Query Auto-Completion. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.  
GGs Rating: A++  
DOI: 10.1145/3397271.3401432
- 2019 Giulio Ermanno Pibiri, Matthias Petri, Alistair Moffat, Fast Dictionary-based Compression for Inverted Indexes. ACM Conference on Web Search and Data Mining (WSDM), pages 9.  
GGs Rating: A+  
DOI: 10.1145/3289600.3290962
- 2017 Giulio Ermanno Pibiri and Rossano Venturini, Efficient Data Structures for Massive N-Gram Datasets. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.  
GGs Rating: A++  
DOI: 10.1145/3077136.3080798
- 2017 Giulio Ermanno Pibiri and Rossano Venturini, Dynamic Elias-Fano Representation. Annual Symposium on Combinatorial Pattern Matching (CPM), pages 14.  
GGs Rating: B  
DOI: 10.4230/LIPIcs.CPM.2017.30

## Posters

- 2021 Raffaele Perego, Giulio Ermanno Pibiri and Rossano Venturini, Compressed Indexes for Fast Search of Semantic Data. IEEE International Conference on Data Engineering (ICDE), pages 2.  
GGs Rating: A++  
DOI: 10.1109/ICDE51399.2021.00248

## Chapters and Theses

- 2019 Giulio Ermanno Pibiri. Space- and Time-Efficient Data Structures for Massive Datasets. Ph.D. Thesis, University of Pisa, 210 pages.
- 2018 Giulio Ermanno Pibiri and Rossano Venturini, Inverted Index Compression. Encyclopedia of Big Data Technologies (EBDT), pages 8.  
DOI: 10.1007/978-3-319-63962-8\_52-1

## Other

- 2021 Giulio Ermanno Pibiri and Roberto Trani, Parallel and External-Memory Construction of Minimal Perfect Hash Functions with PTHash. CoRR, <https://arxiv.org/abs/2106.02350>, pages 12.
- 2020 Giulio Ermanno Pibiri and Rossano Venturini, Succinct Dynamic Ordered Sets with Random Access. CoRR, <https://arxiv.org/abs/2003.11835>, pages 15.
- 2019 Giulio Ermanno Pibiri. On Implementing the Binary Interpolative Coding Algorithm. Tech Report, 8 pages.

## LANGUAGE CERTIFICATIONS

- 2018 — TOEFL iBT in English.  
Grade: 100 (HIGH level)
- 2008 — First Certificate in English (Level B2).  
University of Cambridge, Cambridge, United Kingdom

Data

04/07/2021

Luogo

Prato (PO)