Dichiarazione Sostitutiva di Certificazione (art. 46 DPR n. 445/2000)
Dichiarazione Sostitutiva dell'Atto di Notorietà (art. 47 DPR n. 445/2000)

Il sottoscritto Giulio Ermanno Pibiri

- nato a Bagno a Ripoli (FI) il 13/07/1990
- attualmente residente a Prato (PO) in via Guizzelmi, n. 16, CAP 59100
- reperibile al numero di telefono +39 340 9100050

visto il D.P.R. 28 dicembre 2000, n. 445 concernente "T.U. delle disposizioni legislative e regolamentari in materia di documentazione amministrativa" e successive modifiche ed integrazioni;

vista la Legge 12 novembre 2011, n. 183 ed in particolare l'art. 15 concernente le nuove disposizioni in materia di certificati e dichiarazioni sostituitive;

consapevole che, ai sensi dell'art. 76 del DPR 445/2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono punite ai sensi del Codice penale e delle leggi speciali vigenti in materia, dichiara sotto la propria responsabilità: che quanto dichiarato nel seguente *Curriculum Vitae et Studiorum*, comprensivo delle informazioni sulla produzione scientifica, corrisponde a verità.

Giulio Ermanno Pibiri

Curriculum Vitae et Studiorum

Contact Information

"Ca' Foscari" University of Venice
Department of Environmental Sciences, Informatics and Statistics (DAIS)
Via Torino 155, 30170 Mestre (Venice), Italy

Email	giulioermanno.pibiri@unive.it
Personal page	`https://jermp.github.io`

Personal Information

Place of birth	Bagno a Ripoli (Florence), Italy
Date of birth	13 July 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 & 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 & 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
Colored k-mer Indexing; *Reference Indexing* (RECOMB 2023); *Locality-Preserving Minimal Perfect Hashing for k-mers* (BIOINF 2023); *Compressed and Weighted Dictionaries for k-mers* (BIOINF 2022, WABI 2022, AMB 2023, GBIO 2023); *Time Series Compression* (SPIRE 2021); *Minimal Perfect Hashing* (SIGIR 2021); *Prefix-Sums* (SPE 2020); *Rank/Select Queries* (INFOSYS 2021); *Query Auto-Completion* (SIGIR 2020); *Bitmap Compression* (DCC 2021); *Indexing of Semantic Relations* (TKDE 2020, ICDE 2021); *Indexing and Estimation of Language Models* (SIGIR 2017, TOIS 2019); *Inverted Index Compression* (TOIS 2017, EBDT 2018, WSDM 2019, TKDE 2019, CSUR 2020); *Succinct and Dynamic Ordered Sets of Integers* (CPM 2017).

Research Positions

06/06/2022 –
present
Assistant Professor of Computer Science.
- Department of Environmental Sciences, Informatics and Statistics, "Ca' Foscari" University of Venice, Venice, Italy
- Protocollo n. 52696, 01/06/2022

15/03/2021 –
31/05/2022
Postdoctoral Research Fellow in Computer Science.
- Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy
- 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)
- Reference publications: SIGIR 2021, SPIRE 2021, BIOINF 2022, WABI 2022

01/11/2018 – 28/02/2021	**Postdoctoral Research Fellow in Computer Science**. ○ Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy ○ 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) ○ Reference 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
	○ Thesis: *Space- and Time-Efficient Data Structures for Massive Datasets*
	○ Supervisor: Rossano Venturini (https://rossanoventurini.github.io)
	○ Reference publications: TKDE 2019, TOIS 2019, WSDM 2019, SIGIR 2017, CPM 2017, TOIS 2017
	
	Part of the research was conducted **abroad** (6 months):
	○ 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: 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: INFOSYS 2021

Research Visits

19/10/2022 – 21/10/2022	Inria, Rennes, France. Visiting dr. Pierre Peterlongo and dr. Karel Brinda.
10/05/2022 – 12/05/2022	The University of Lille, CNRS, CRIStAL Laboratory, Lille, France. Visiting dr. Antoine Limasset and dr. Camille Marchet.
01/05/2018 – 01/10/2018	The University of Melbourne, School of Computing and Information Systems, Melbourne, Australia. Visiting prof. Alistair Moffat and dr. Matthias Petri.
01/04/2018 – 30/04/2018	RIKEN Advanced Intelligence Project (AIP), Tokyo, Japan. Visiting dr. Yasuo Tabei and dr. Shunsuke Kanda.

Projects

01/09/2021 – 31/05/2021	**Italian PON project OK-INSAID**. ○ Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy ○ Role: Investigator
15/03/2021 – 31/05/2021	**European project ACCORDION**. ○ Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy ○ Role: Investigator
01/11/2018 – 28/02/2021	**European project BIGDATAGRAPES**. ○ Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy ○ Role: *Task Leader* for the task 3.3 "Big Data Indexing"
01/06/2017 – 31/10/2018	**European project LIGA**. ○ Institute of Science and Information Technologies "A. Faedo" (ISTI), National Research Council of Italy (CNR), Pisa, Italy ○ Role: Principal Developer

Teaching Activity

A.Y. 2022/2023   Teacher for Introduction to Algorithms (3h), Invited lecture, IIS A. Pacinotti, Mestre (Venice), Italy

A.Y. 2022/2023   Teacher for KNOWLEDGE, INTERACTION AND INTELLIGENT SYSTEMS-2 (2 CFU), PhD Program in Computer Science, Ca' Foscari University, Italy

A.Y. 2022/2023   Teacher for COMPUTER SCIENCE I - MOD. 1 (code CT0569, 6 CFU), Bachelor Degree in Ingegneria Fisica, Ca' Foscari University, Italy

A.Y. 2022/2023   Teacher for COMPUTER SCIENCE FOR CULTURAL HERITAGE (code CT0612, 6 CFU), Bachelor Degree in SCIENZE E TECNOLOGIE PER I BENI CULTURALI, Ca' Foscari University, Italy

A.Y. 2022/2023   Teacher for PROGRAMMING AND LABORATORY-2 (code CT0442, 6 CFU), Bachelor Degree in INFORMATICA, Ca' Foscari University, Italy

A.Y. 2021/2022   Teacher for *Theory and Practice of Data Compression* (5 CFU), PhD Program in Ingegneria dell'Informazione, University of Pisa, Italy

A.Y. 2019/2020   Teacher for *Algorithmics and Laboratory - Corso B* (code 008AA, 3 CFU), Bachelor Degree in Computer Science, University of Pisa, Italy

A.Y. 2018/2019   Assistant for *Algorithmics and Laboratory - Corso A* (code 008AA, 3 CFU), Bachelor Degree in Computer Science, University of Pisa, Italy

A.Y. 2017/2018   Assistant for *Competitive Programming and Contests* (code 645AA, 6 hours), Master Degree in Computer Science, University of Pisa, Italy

A.Y. 2016/2017   Assistant for *Competitive Programming and Contests* (code 645AA, 6 hours), Master Degree in Computer Science, University of Pisa, Italy

A.Y. 2015/2016   Teacher for *Algorithmics and Laboratory - Corso di recupero* (code 008AA, 3 CFU), Bachelor Degree in Computer Science, University of Pisa, Italy

A.Y. 2015/2016   Assistant for *Algorithmics and Laboratory - Corso A* (code 008AA, 3 CFU), Bachelor Degree in Computer Science, University of Pisa, Italy

Awards and Grants

2022   *Young Researcher Award* for the year 2021, issued by ISTI-CNR.

2021   *Young Researcher Award* for the year 2020, issued by ISTI-CNR.

2020   *Young Researcher Award* for the year 2019, 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* for the 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).

Organizing Committees

2023   Chair of the Workshop on Compression, Text, and Algorithms (WCTA), co-located with SPIRE 2023.

2020   The 28-th edition of the Annual European Symposium on Algorithms (ESA 2020).

2019   The 30-th edition of the International Symposium on Combinatorial Pattern Matching (CPM 2019).

2017   The 24-th International Symposium on String Processing and Information Retrieval (SPIRE 2017).

2016   The 39-th ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016). (As student volunteer.)

Program Committees

2023	The 31-st European Symposium on Algorithms (ESA 2023 – Track B).
2023	The 13-rd RECOMB Satellite Conference on Biological Sequence Analysis (RECOMB-SEQ 2023).
2023	The 13-rd International Symposium on Algorithms and Complexity (CIAC 2023).
2023	The 45-th European Conference on Information Retrieval (ECIR 2023).
2023	The 16-th International ACM Conference on Web Search and Data Mining (WSDM 2023).
2022	The 45-th ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022).
2022	The 44-th European Conference on Information Retrieval (ECIR 2022).
2022	The 15-th International ACM Conference on Web Search and Data Mining (WSDM 2022).
2021	The 15-th IEEE International Conference on Application of Information and Communication Technologies (AICT 2021).
2021	The 30-th ACM International Conference on Information and Knowledge Management (CIKM 2021).
2021	The 44-th ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2021).
2021	The 43-rd European Conference on Information Retrieval (ECIR 2021).
2021	The 14-th International ACM Conference on Web Search and Data Mining (WSDM 2021).
2020	The 29-th ACM International Conference on Information and Knowledge Management (CIKM 2020).
2020	The 43-rd ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020).
2019	The 42-nd ACM International SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).

Reviewing Activity

2016 – present	I am or have been an anonymous reviewer for the following 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
- SPIRE – String Processing and Information Retrieval
- ISAAC – International Symposium on Algorithms and Computation
- RECOMB – International Conference on Research in Computational Molecular Biology
- RECOMB-SEQ – Satellite Conference on Biological Sequence Analysis.
- CIAC – International Symposium on Algorithms and Complexity

Journals

- TALG – ACM Transactions on Algorithms
- Bioinformatics
- INFOSYS – Information Systems
- SPE – Software: Practice and Experience
- JEA – Journal of Experimental Algorithmics
- Algorithmica
- MDPI Algorithms

Talks

Invited Keynote

20/10/2022 *Modular reference indexing with the de Bruijn graph: overview and challenges*. Workshop on Indexing Omic Sequences. Rennes, France.

10/05/2022 *On Weighted K-Mer Dictionaries*. TUDASTIC 2022 (TUtorials on DAta Structures for Text Indexation and Compression). Lille, France.

Conferences

21/03/2023 *Spectrum Preserving Tilings Enable Sparse and Modular Reference Indexing*. DSB 2023, Delft, Netherlands.

05/09/2022 *On Weighted K-Mer Dictionaries*. WABI 2022, Potsdam, Germany.

14/06/2022 *PTHash: Revisiting FCH Minimal Perfect Hashing*. DSB 2022, Düsseldorf, Germany.

20/05/2022 *Sparse and Skew Hashing of K-Mers*. RECOMB-seq 2022. La Jolla, California, USA.

04/10/2021 *TSXor: A Simple Time-Series Compression Algorithm*. SPIRE 2021. Lille, France (Virtual event).

14/09/2021 *PTHash: Revisiting FCH Minimal Perfect Hashing*. IIR 2021. Polytechnic University of Bari, Bari (Virtual event).

07/2021 *PTHash: Revisiting FCH Minimal Perfect Hashing*. SIGIR 2021. Montreal, Canada (Virtual event).

04/2021 *Compressed Indexes for Fast Search of Semantic Data*. ICDE 2021. Chania, Greece (Virtual event).

03/2021 *Fast and Compact Set Intersection through Recursive Universe Partitioning*. DCC 2021. Snow Bird, USA (Virtual event).

27/07/2020 *Efficient and Effective Query Auto-Completion*. SIGIR 2020. Xi'An, China (Virtual event).

17/09/2019 *Compressed Indexes for Fast Search of Semantic Data*. IIR 2019. Department of Information Engineering, Padova, Italy.

12/02/2019 *Fast Dictionary-based Compression for Inverted Indexes*. WSDM 2019. Melbourne Exhibition Center, Melbourne, Australia.

10/08/2017 *Efficient Data Structures for Massive N-Gram Datasets*. SIGIR 2017. Keio Plaza Hotel, Tokyo, Japan.

06/07/2017 *Dynamic Elias-Fano Representation*. CPM 2017. University Library of Warsaw, Warsaw, Poland.

06/06/2017 *Efficient Data Structures for Massive N-Gram Datasets*. IIR 2017. Università della Szizzera Italiana, Lugano, Switzerland.

Seminars

19/05/2022 *Sparse and Skew Hashing of K-Mers*. ISTI-CNR, Pisa, Italy (Virtual event).

22/12/2021	*Minimal Perfect Hashing and K-Mer String Dictionaries*. "Ca' Foscari" University of Venice, Venice, Italy (Virtual event).
16/11/2021	*PTHash: Revisiting FCH Minimal Perfect Hashing*. ISTI-CNR, Pisa, Italy (Virtual event).
04/03/2021	*Efficiency for Real-World Applications*. ISTI-CNR, Pisa, Italy (Virtual event).
07/06/2019	*Ordered Set Problems*. ISTI-CNR, Pisa, Italy.
01/02/2019	*Indexing Compressed Data for Fast Retrieval*. University of Pisa, Pisa, Italy.
29/10/2018	*Effective Web Graph Representations*. University of Pisa, Pisa, Italy.
17/05/2018	*On Optimally Partitioning Variable-Byte Index Data*. RMIT University, Melbourne, Australia.
10/04/2018	*Elias-Fano Encoding: a powerful tool for data structure design*. RIKEN AIP, Tokyo, Japan.
21/06/2016	*Elias-Fano Encoding: succinct representation of monotone integer sequences with search operations*. University of Pisa, Pisa, Italy.

Publications

Journal Papers

GB 2023   Sebastian Schmidt, Shahbaz Khan, Jarno Alanko, Giulio Ermanno Pibiri, Alexandru I. Tomescu, *Matchtigs: minimum plain text representation of k-mer sets*. 2023. Genome Biology (GBIO), pages 33.
Scimago Rating: **Q1**
DOI:
ISSN: 1474-7596

BIOINF 2023   Giulio Ermanno Pibiri, Yoshihiro Shibuya, and Antoine Limasset, *Locality-Preserving Minimal Perfect Hashing of k-mers*. 2022. Bioinformatics (BIOINF), pages 9.
Scimago Rating: **Q1**
DOI:
ISSN: 1367-4803

AMB 2022   Giulio Ermanno Pibiri, *On Weighted k-mer Dictionaries*. 2023. Algorithms for Molecular Biology (AMB), pages 25.
Scimago Rating: **Q2**
DOI:
ISSN: 1748-7188

BIOINF 2022   Giulio Ermanno Pibiri, *Sparse and Skew Hashing of k-mers*. 2022. Bioinformatics (BIOINF), pages 9.
Scimago Rating: **Q1**
DOI: 10.1093/bioinformatics/btac245
ISSN: 1367-4803

INFOSYS 2021   Giulio Ermanno Pibiri and Shunsuke Kanda, *Rank/Select Queries over Mutable Bitmaps*. 2021. Information Systems (INFOSYS), pages 15.
Scimago Rating: **Q2**
DOI: 10.1016/j.is.2021.101756
ISSN: 0306-4379

CSUR 2020   Giulio Ermanno Pibiri and Rossano Venturini, *Techniques for Inverted Index Compression*. 2020. ACM Computing Surveys (CSUR), pages 36.
Scimago Rating: **Q1**
DOI: 10.1145/3415148
ISSN: 0360-0300

SPE 2020   Giulio Ermanno Pibiri and Rossano Venturini, *Practical Trade-Offs for the Prefix-Sum Problem*. 2020. Software: Practice and Experience (SPE), pages 29.
Scimago Rating: **Q2**
DOI: 10.1002/spe.2918
ISSN: 0038-0644

TKDE 2020   Raffaele Perego, Giulio Ermanno Pibiri and Rossano Venturini, *Compressed Indexes for Fast Search of Semantic Data*. 2020. IEEE Transactions on Knowledge and Data Engineering (TKDE), pages 12.
Scimago Rating: **Q1**
DOI: 10.1109/TKDE.2020.2966609
ISSN: 1041-4347

TKDE 2019   Giulio Ermanno Pibiri and Rossano Venturini, *On Optimally Partitioning Variable-Byte Codes*. 2019. IEEE Transactions on Knowledge and Data Engineering (TKDE), pages 12.
Scimago Rating: **Q1**
DOI: 10.1109/TKDE.2019.2911288
ISSN: 1041-4347

TOIS 2019   Giulio Ermanno Pibiri and Rossano Venturini, *Handling Massive N-Gram Datasets Efficiently*. 2019. ACM Transactions on Information Systems (TOIS), pages 41.
Scimago Rating: **Q1**
DOI: 10.1145/3302913
ISSN: 1046-8188

TOIS 2017   Giulio Ermanno Pibiri and Rossano Venturini, *Clustered Elias-Fano Indexes*. 2017. ACM Transactions on Information Systems (TOIS), volume 2, pages 33.
Scimago Rating: **Q1**
DOI: 10.1145/3052773
ISSN: 1046-8188

Conference Papers

RECOMB 2023   Jason Fan, Jamshed Khan, Giulio Ermanno Pibiri, and Rob Patro, *Spectrum preserving tilings enable sparse and modular reference indexing*. 2023. International Conference on Research in Computational Molecular Biology (RECOMB), 20 pages.
GGS Rating: **A++**
DOI: 10.1007/978-3-031-29119-7_2
ISBN: 9783031291197

WABI 2022   Giulio Ermanno Pibiri, *On Weighted k-mer Dictionaries*. 2022. International Workshop on Algorithms in Bioinformatics (WABI), 20 pages.
GGS Rating: B
DOI: 10.4230/LIPIcs.WABI.2022.9
ISBN: 9783959772433

SPIRE 2021   Andrea Bruno, Franco Maria Nardini, Giulio Ermanno Pibiri, Roberto Trani, and Rossano Venturini, *TSXor: A Simple Time Series Compression Algorithm*. 2021. International Symposium on String Processing and Information Retrieval (SPIRE), 8 pages.
GGS Rating: B
DOI: 10.1007/978-3-030-86692-1_18
ISBN: 9783030866921

SIGIR 2021   Giulio Ermanno Pibiri and Roberto Trani, *PTHash: Revisiting FCH Minimal Perfect Hashing*. 2021. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.
GGS Rating: **A++**
DOI: 10.1145/3404835.3462849
ISBN: 9781450380379

DCC 2021   Giulio Ermanno Pibiri, *Fast and Compact Set Intersection through Recursive Universe Partitioning*. 2021. IEEE Data Compression Conference (DCC), pages 10.
GGS Rating: **A**-
DOI: 10.1109/DCC50243.2021.00037
ISBN: 9781665403337

SIGIR 2020   Simon Gog, Giulio Ermanno Pibiri and Rossano Venturini, *Efficient and Effective Query Auto-Completion*. 2020. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.

WSDM 2019   Giulio Ermanno Pibiri, Matthias Petri, Alistair Moffat, *Fast Dictionary-based Compression for Inverted Indexes*. 2019. ACM Conference on Web Search and Data Mining (WSDM), pages 9.

SIGIR 2017   Giulio Ermanno Pibiri and Rossano Venturini, *Efficient Data Structures for Massive N-Gram Datasets*. 2017. ACM Conference on Research and Development in Information Retrieval (SIGIR), pages 10.

CPM 2017   Giulio Ermanno Pibiri and Rossano Venturini, *Dynamic Elias-Fano Representation*. 2017. Annual Symposium on Combinatorial Pattern Matching (CPM), pages 14.

Posters

ICDE 2021   Raffaele Perego, Giulio Ermanno Pibiri and Rossano Venturini, *Compressed Indexes for Fast Search of Semantic Data*. 2021. IEEE International Conference on Data Engineering (ICDE), pages 2.

PhD Thesis

2019   Giulio Ermanno Pibiri. *Space- and Time-Efficient Data Structures for Massive Datasets*. 2019. Ph.D. Thesis, University of Pisa, 210 pages.

Chapters

EBDT 2018   Giulio Ermanno Pibiri and Rossano Venturini, *Inverted Index Compression*. 2018. Encyclopedia of Big Data Technologies (EBDT), pages 8.

Pre-Prints

2023   Jason Fan, Noor Pratap Singh, Jamshed Khan, Giulio Ermanno Pibiri, Rob Patro, *Fulgor: A fast and compact k-mer index for large-scale matching and color queries*. 2023. bioRxiv, `https://www.biorxiv.org/content/10.1101/2023.05.09.539895v2`, pages 21.

2023   Stefano Calzavara, Lorenzo Cazzaro, Giulio Ermanno Pibiri, and Nicola Prezza, *Verifiable Learning for Robust Tree Ensembles*. 2023. ArXiv, `https://arxiv.org/abs/2305.03626`, pages 17.

2021   Giulio Ermanno Pibiri and Roberto Trani, *Parallel and External-Memory Construction of Minimal Perfect Hash Functions with PTHash*. 2021. ArXiv, `https://arxiv.org/abs/2106.02350`, pages 12.

2020   Giulio Ermanno Pibiri and Rossano Venturini, *Succinct Dynamic Ordered Sets with Random Access*. 2020. ArXiv, `https://arxiv.org/abs/2003.11835`, pages 15.

2019   Giulio Ermanno Pibiri. *On Implementing the Binary Interpolative Coding Algorithm*. 2019. Tech Report, 8 pages.

Software

GitHub profile   All software is open-source and available from `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), Locality-Preserving Minimal Perfect Hash Functions (BIOINF 2023)
- k-mer Dictionaries (BIOINF 2022, WABI 2022, AMB 2023, GBIO 2023)
- Colored k-mer indexes (bioRxiv 2023)

A more detailed list follows below.

fulgor   Fulgor: A fast and compact k-mer index for large-scale matching and color queries.
Reference publication: bioRxiv 2023.

sshash   SSHash: A compressed, weighted, associative, exact dictionary for k-mers. (See also the membership-only version **sshash-lite**.)
Reference publication: BIOINF 2022, WABI 2022, AMB 2023, GBIO 2023.

lphash   LPHash: Fast and compact locality-preserving minimal perfect hashing for k-mer sets.
Reference publication: BIOINF 2023.

pthash   PTHash: Fast and compact minimal perfect hash functions.
Reference publications: SIGIR 2021, ArXiv 2106.02350.

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

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.

autocomplete   Efficienct and effective autocompletion framework, based on forward/inverted indexes, succinct RMQ, and string dictionaries (Front-Coding and tries).
Reference publication: SIGIR 2020.

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.

interp	An efficient implementation of the Binary Interpolative Coding algorithm.
s_indexes	Compressed bitmap indexes that support fast intersection and union. Reference publication: DCC 2021.
rdf_indexes	Trie-based indexes for semantic data like RDF triples. Reference publication: TKDE 2020.
dint	DINT: fast and compact dictionary-based decoder for inverted lists. Reference publication: WSDM 2019.
opt_vbyte	Optimal partitioning of inverted lists compressed using binary vectors and point-wise encoders, like Variable-Byte. Reference publication: TKDE 2019.
tongrams	Fast language model queries and estimation in compressed space. Reference publications: SIGIR 2017, TOIS 2019.
clustered_indexes	Clustered Elias-Fano inverted indexes. Reference publication: TOIS 2017.

Miscellanea

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.
cmd_line	Command line parser for C++17. It offers all handy features in just 150 lines of code.
mm_file	A self-contained, header-only, implementation of memory-mapped files in C++ for both reading and writing.

Languages

Italian	**Native**	*CEFR level: C2*
English	**Fluent**	*CEFR level: C1*
2018	**TOEFL iBT in English**. 100 (HIGH level)	
2008	**First Certificate in English (Level B2)**. University of Cambridge, Cambridge, United Kingdom	