

Curriculum Vitae

Paolo Ferragina
Università di Pisa, Italy
paolo.ferragina@unipi.it

I am Full Professor of Algorithms at the University of Pisa and the current President of the Computer Science PhD joint program from the University of Pisa, Florence and Siena. I founded and lead the Acube Lab (*Advanced Algorithms and Applications*) of the University of Pisa where we design algorithms for Big Data collaborating with major companies worldwide; such as Google, Bloomberg, Tiscali, Yahoo!, ST Microelectronics, ENEL, CERVED, etc.

Previous Institutional Roles. From 2010 to 2016 I was Vice-Rector on “Applied Research and Innovation” at the University of Pisa. The two most significant initiatives during my term were: (1) the MIT-UniPI collaboration which led to the development of 30+ projects jointly pursued by researchers of MIT and UniPI; (2) the *PhD Plus* program: a series of lectures on research valorization and entrepreneurship for PhDs, Master students, and Faculties of the University of Pisa. This program got 600+ attendees and originated many start-ups that won (inter-)national competitions and raised funding worldwide. The PhD Plus program was shortlisted in the *QS Re-imagine Education Award 2016* (Philadelphia, PA).

In the last six years I was also the President of the IT Center of the University of Pisa, which has been recognized by Dell and Intel as a worldwide competence center on Cloud and HPC, and recently appointed as Transform Data Center by Microsoft. Previously, I was scientific coordinator of the *Signum* research center on Digital Humanities at the Scuola Normale Superiore in Pisa.

Education and Research interests. I got the PhD in Computer Science from the University of Pisa (1996), and the Post-doc from the Max-Planck Institut für Informatik (1998). Then I joined University of Pisa, where I am full professor since 2007. My professorship was initially sponsored by Yahoo! with a Faculty Award from 2006 to 2010 (the first in Europe at that time) because of my scientific achievements on the topic of Compressed and I/O-efficient data structures.

I have spent research periods mainly outside academia: IBM Research (Rome), AT&T Shannon Lab (Florham Park, NJ), Yahoo! Research (Barcelona), Google (New York and Zurich), University of North Texas, Max Planck Institut fuer Informatik (Saarbrcken, DE), and Courant Institute (New York).

My research has been mainly devoted to the design, analysis and experimentation of algorithms and data structures for the compressed storage, mining and search of information in large amounts of data, mainly in the form of texts and (labeled) graphs. My results draw from the fields of String Matching, Information Theory, Data Compression, Graph Algorithms and, starting from 2006, also Web search and mining. I can describe my research by distinguishing four main periods/challenges:

- During the years of my PhD and post-doc, I investigated the design of I/O-efficient data structures for String Matching problems. The main result of my Phd Thesis was the String B-tree [ref. 1] mentioned nowadays in several US patents and published first in the *Procs of ACM STOC (1994)* and then in the *Journal of the ACM (2005)*. My thesis got the “EATCS Doctoral Dissertation Thesis Award” (1997) and the “Philip Morris Award on Science and Technology” (1997); moreover, it was ranked 4th in the “1997 ACM PhD Thesis” competition.
- In the year 2000, following a research period at AT&T Shannon Lab, I turned my attention to Information Theory. My main result has been the FM-index which is the first data structure supporting efficient substring searches over arbitrary data compressed up their k -th order entropy [ref. 3]. Researchers have recognized the FM-index as the first compressed version of the well-known

suffix tree and suffix array data structures, as celebrated in the paper on the “40 years of Suffix Trees (Comm. ACM, 2016). More importantly, the FM-index initiated the field of *compressed data structures*, nowadays the topic of thousands of publications (see e.g. the book by G. Navarro published by Cambridge University Press in 2016) and it plays a key algorithmic role in the two most famous genome-alignment tools, namely Bowtie and BWA, which are used by bio-researchers all over the world as witnessed by their 32k+ citations on Google Scholar. More recently, I was able to mix optimality in (compressed) space with cache obliviousness in I/Os [ref. 8], and to extend those ideas to labelled trees [ref. 4], graphs [ref. 11] and Web search [ref. 7].

- From 2005 I started working on Data Compression and got two main results. The first one is the *Compression Booster* [ref. 5], an inherently combinatorial technique that turns any *memoryless* compressor into a *kth-order entropy* compressor based on a suitable combination of the Burrows-Wheeler Transform and Suffix Trees. The second contribution [ref. 6,10] is the so called *bi-criteria compressor* that, given as input some data and a constant t , computes the Lempel-Ziv parsing of that data which is minimum in space and can be decompressed in time t (or, vice versa). Experimentally, the bi-criteria compressor dominates highly engineered competitors, such as Snappy, LZ4, bzip and LZMA. This research got a Google Faculty Award in 2016.
- Since 2010 I have been interested in *Semantic Text Annotators* [ref. 9]. These algorithms are able to detect and annotate sequences of terms with unambiguous and pertinent entities drawn from a catalog (typically, Wikipedia). I lead the development of *TagMe*, which is currently the best known open-source annotator in terms of efficiency and efficacy [ref. 12]. Some of the algorithmic ideas developed in the previous years have been used to make TagMe extremely efficient in time and space and successfully used in applications: such as news clustering [ACM WSDM '12] and classification [ECIR '12], analysis of hashtags in tweets [ICWSM '15], and entity salience and relatedness [NLDB '17, ACM CIKM '17]. TagMe has been released under Apache 2.0 License and its API interface has been queried more than 600mln times from academics and industries. On this research I got two Google Faculty Awards (2010 and 2013), and a “Working Capital Award” from Telecom Italia (2010). Recently, TagMe has been used to design SMAPH a system for entity-annotation of Web queries that won the ERD 2014 Short Track Challenge at ACM SIGIR 2014 [ref. 13].¹

Patents and Awards. I am co-inventor of 3 US Patents (owned by Lucent, University of Pisa and Rutgers, Yahoo!), and 3 more patents are pending (2 owned by Yahoo! and 1 by NYU). My research also got many international awards: “Best Land Transportation Paper Award” from IEEE Vehicular Technology Society (1995); “EATCS Doctoral Dissertation Thesis Award” (1997); “Philip Morris Award on Science and Technology” (1997); “Research Capital award” from the University of Pisa (2002); Yahoo! faculty award (2006-2010); Working Capital Award from Telecom Italia (2010); three Google research awards (2010, 2013 and 2016); and a Bloomberg Data Science research grant (2017).

Conferences keynotes and co-chairs. I have been plenary speaker at the conferences CPM '04, SPIRE '05, ESA/ALGO 2010, SISAP 2011, and ECIR Industrial-Track 2012, DFG Priority Program 1307 “Algorithm Engineering” (Germany) and the DIITET 2015 National Conference of CNR (Pisa).

I've been co-chair of international conference on FUN with Algorithms (2004), DIMACS Workshop on the Burrows-Wheeler Transform (2004), Symposium on String Processing and Information Retrieval (SPIRE 2006), Symposium on Combinatorial Pattern Matching (CPM 2008), European Symposium on Algorithms (ESA 2012), ACM Conference on Web Search and Data Mining (WSDM 2013). I have been a member of the steering committees of the conferences SPIRE (from 2006 to 2009), ESA (from 2012 to 2014), and of the Lipari PhD School su “Computational Complex and Social Systems” (from 2016).

Of course, I have served as PC member of many Theoretical Computer Science conferences, specifically in the field of Algorithmics (i.e. ESA, ICALP, CPM, etc.) and IR (i.e. WWW, WSDM, SPIRE, etc.).

Journals and Books. Since 2011 I'm in the Editorial Board of the Journal on Graph Algorithms and Applications (JGAA), and I have served as (co)editor of 5 special issues in Theory of Computing Systems

¹See the Google's Research post: <https://plus.google.com/+ResearchatGoogle/posts/aqU7GrhXDLy>

(2006), Theoretical Computer Science (2007 and 2009), Information Retrieval (2008), Algorithmica (2014). I was one of the Area Editors of the Encyclopedia of Algorithms (Springer, Editor Ming-Yang Kao).

I (co-)authored more than 150 (refereed) publications in journals — such as 5 JACM, 2 SICOMP, 4 ACM Trans on Algorithms, 7 Algorithmica, 6 TCS, 2 ACM J. on Exp. Algorithmics, 2 Information and Computation, etc.. — and conferences — such as 3 ACM STOC, 5 IEEE FOCS, 7 ACM-SIAM SODA, 13 ESA, 6 WWW, etc.. I have also co-authored two Italian books: one on Cryptography (Bollati Boringhieri, 2001 and 2007; now UniPI Press, 2015), and one on Computational Thinking (Il Mulino, 2017, and Springer, 2018 – to appear). I have also contributed to several Encyclopedia’s entries by Springer (on Algorithms, on DataBases, and on BigData Technologies) and various book chapters: e.g. “String search in external memory: Algorithms and data structures” in the “Handbook of Computational Molecular Biology” (CRC Press, 2005), “Web Search” in the book “On the power of algorithms” (Springer, 2013), and “Computational Biology” in the “Handbook of Data Structures” (Springer, 2017).

My H-index on Google Scholar is 38 (with 7k+ citations, at December 2017). For more details have a look at: <http://pages.di.unipi.it/ferragina>, DBLP and Google Scholar.

Most significant and pertinent publications

1. P. Ferragina, R. Grossi. The String B-tree: a new data structure for string search in external memory and its applications, *Journal of the ACM*, 46(2), 1999. Its conference version appeared in the *Procs ACM STOC 2005*.
2. M. Farach, P. Ferragina, S. Muthukrishnan. On the sorting complexity of suffix tree construction. *Journal of the ACM*, 47(6), 2000. Its conference version appeared in the *Procs IEEE FOCS 1998*.
3. P. Ferragina, G. Manzini. Indexing compressed texts. *Journal of the ACM*, 52(4), 2005. Its conference version appeared in the *Procs IEEE FOCS 2000*.
4. P. Ferragina, F. Luccio, G. Manzini, S. Muthukrishnan. Compressing and indexing labeled trees, with applications. *Journal of the ACM*, 57(1), 2009. Its conference version appeared in the *Procs IEEE FOCS 2005* and *Procs WWW 2006*.
This result is the subject of the US Patent no. US8156156 (10 apr '12).
5. P. Ferragina, R. Giancarlo, G. Manzini, M. Sciortino. Compression boosting in optimal linear time. *Journal of the ACM*, 52(4), 2005. Its conference version appeared in the *Procs ACM-SIAM SODA 2004*.
6. P. Ferragina, I. Nitto, R. Venturini. On the bit-complexity of Lempel-Ziv compression. *SIAM Journal on Computing*, 42(4), 2013. Its conference version appeared in the *Procs ACM-SIAM SODA 2009*.
7. P. Ferragina, R. Venturini. The compressed Permuterm index. *ACM Transactions on Algorithms* 7(1), 2010. Its conference version appeared in the *Procs ACM SIGIR 2007*.
This result is the subject of the US Patent Application 20090063465 (5 mar '09).
8. P. Ferragina, R. Venturini. Compressed cache-oblivious String B-Tree. *ACM Transactions on Algorithms*, 12(4), 2016. Its conference version appeared in the *Procs ESA 2013*.
9. P. Ferragina, U. Scaiella. Fast and accurate annotation of short texts with Wikipedia pages. *IEEE Software*, 29(1), 2012. Its conference version appeared in the *Procs ACM CIKM 2010*.
10. A. Farruggia, P. Ferragina, A. Frangioni, R. Venturini. Bicriteria data compression. In *Procs ACM-SIAM SODA*, 2014. Its experimental results have been presented in *European Symposium on Algorithms (ESA)*, 2014.
11. P. Ferragina, F. Piccinno, R. Venturini. Compressed indexes for string-searching in labeled graphs. In *Procs WWW*, 2015.
12. R. Usbeck, M. Röder, P. Ferragina *at al.*. GERBIL – General entity annotator benchmark. In *Procs World Wide Web Conference (WWW)*, 2015.
13. M. Cornolti, P. Ferragina, M. Ciaramita, S. Rued, H. Schutze. A piggyback system for joint entity mention detection and linking in Web queries. In *Procs WWW*, 2016. (Winner of the Short-Track competition at the Workshop ERD, hosted by ACM SIGIR 2014.)

Full list of publications

Journals

- [1] N. Burgess, S. Di Zenzo, P. Ferragina, M. Notturmo Granieri. The generalization of a constructive algorithm in pattern classification problems. *International Journal of Neural Systems: Supplementary Issue on the Second Workshop on Neural Networks– from Biology to High Energy Physics*, World Scientific Publishing, vol. 3, 65-70, 1992.
- [2] S. Di Zenzo, N. Burgess, P. Ferragina, M. Notturmo Granieri. Recognition by constructive neural algorithms. *Pattern Recognition Letters*, vol. 14(12), 997-1007, 1993.
- [3] C. Calabrò, P. Ferragina, M. Notturmo Granieri. Recognition of hand-written rotated digits by neural networks. *Machine Vision and Applications*, Springer Verlag, vol. 8, 351–358, 1995.
- [4] P. Comelli, P. Ferragina, M. Notturmo Granieri, F. Stabile. Optical recognition of motor vehicle license plates. *IEEE Transactions on Vehicular Technology*, vol. 44(4), 790–799, 1995.
- [5] P. Ferragina. Static and dynamic parallel computation of connected components. *Information Processing Letters*, vol. 50(2), 63–68, 1994.
- [6] P. Ferragina. A technique to speed up parallel fully dynamic algorithms for MST. *Journal of Parallel and Distributed Computing*, vol. 31(2), 181–189, 1995.
- [7] P. Ferragina, F. Luccio. Three techniques for parallel maintenance of a minimum spanning tree under batch of updates. *Parallel Processing Letters*, vol. 6(2), 213–222, 1996.
- [8] P. Ferragina. Dynamic Text Indexing under string updates. *Journal of Algorithms*, vol. 22(2), 296–328, 1997.
- [9] P. Ferragina, R. Grossi. Optimal on-line search and sublinear time update in string matching. *SIAM Journal on Computing*, vol. 27(3), 713–736, 1998.
- [10] P. Ferragina, R. Grossi, M. Montangero. A note on updating suffix tree labels. *Theoretical Computer Science*, vol. 201(1–2), 249–262, 1998.
- [11] P. Ferragina, F. Luccio. Dynamic Dictionary Matching in External Memory. *Information and Computation*, vol. 146(2), 85–99, 1998.
- [12] A. Crauser, P. Ferragina, K. Mehlhorn, U. Meyer, E. Ramos. An I/O-optimal randomized algorithm for the segment intersections problem. *Volume on External Memory Algorithms and/or Visualization*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, James Abello and Jeffrey S. Vitter Eds, American Mathematical Society, 1998.
- [13] P. Ferragina, F. Luccio. String search in coarse-grained parallel computers. *Algorithmica: Special issue on Coarse-Grained Parallel Computers*, Editor Frank Dehne, vol. 24(3), 177–194, 1999.
- [14] P. Ferragina, R. Grossi. Improved Dynamic Text Indexing. *Journal of Algorithms*, vol. 31(2), pp. 291–319, 1999.
- [15] P. Ferragina, R. Grossi. The String B-Tree: A New Data Structure for String Search in External Memory and its Applications. *Journal of the ACM*, vol. 46(2), pp. 236–280, March 1999.
- [16] S. K. Das, P. Ferragina. An EREW PRAM algorithm for updating minimum spanning trees. *Parallel Processing Letters*, 9(1), pp. 111–122, 1999.
- [17] M. Farach, P. Ferragina, S. Muthukrishnan. On the sorting complexity of suffix tree construction. *Journal of the ACM*, vol. 47(6), pp. 987–1011, November 2000.

- [18] K. Brengel, A. Crauser, P. Ferragina, U. Meyer. An Experimental Study of Priority Queues in External Memory. *ACM Journal on Experimental Algorithmics* (Special Issue WAE '99), vol. 5, art. 17, 2000.
- [19] A. Crauser, P. Ferragina, K. Mehlhorn, U. Meyer, E. Ramos. Randomized external-memory algorithms for some geometric problems. *International Journal on Computational Geometry and Applications* (Special issue on ACM SoGC '98), 11(3): 305–339, 2001.
- [20] P. Ferragina, G. Manzini. An experimental study of a compressed index. Articolo invitato in *Information Sciences: special issue on Dictionary Based Compression*, vol. 135(1-2), pp. 13–28, 2001.
- [21] A. Crauser, P. Ferragina. A theoretical and experimental study on the construction of suffix arrays in external memory. *Algorithmica*, 32(1):1–35, 2002.
- [22] P. Ferragina, N. Koudas, S. Muthukrishnan, D. Srivastava. Two-dimensional substring indexing. *Journal of Computer and System Sciences: Special Issue on ACM PODS '01*, 66(4):763–774, 2003.
- [23] N. Pisanti, R. Marangoni, P. Ferragina, A. Frangioni, A. Savona, C. Pisanelli, F. Luccio. PaTre: A method for Paralogy Trees construction. *Journal of Computational Biology*, 10(5): 791–802, 2003.
- [24] G. Manzini, P. Ferragina. Engineering a lightweight suffix array construction algorithm. *Algorithmica*, 40(1):33–50, 2004.
- [25] P. Ferragina, G. Manzini. Indexing compressed texts. *Journal of the ACM*, 52(4):552–581, 2005.
- [26] P. Ferragina, R. Giancarlo, G. Manzini, M. Sciortino. Compression boosting in optimal linear time. *Journal of the ACM*, 52(4):688–713, 2005.
- [27] C. Corsi, P. Ferragina, R. Marangoni. The BioPrompt-box: an ontology-based clustering tool for searching in biological databases. *BMC Bioinformatics*, 8(Suppl 1), 2007.
- [28] P. Ferragina, R. Venturini. On a simple storage scheme for strings achieving entropy bounds. *Theoretical Computer Science*, 372:115–121, 2007.
- [29] V. Ciriani, P. Ferragina, F. Luccio, S. Muthukrishnan. A data structure for a sequence of string accesses in external memory. *ACM Transactions on Algorithms*, 3(1), 2007.
- [30] P. Ferragina, A. Gulli. Snaket: A personalized search-result clustering engine. *The European Journal for the Informatics Professionals*, Upgrade–digital journal of CEPIS (ISSN 1684-5285), 8(1), 19–26, 2007.
- [31] P. Ferragina, G. Manzini, V. Mäkinen, G. Navarro. Compressed Representations of sequences and full-text indexes. *ACM Transactions on Algorithms*, 3(2), 2007.
- [32] P. Ferragina, R. Giancarlo, V. Greco, G. Manzini, G. Valiente. Kolmogorov Complexity, Information Theory and Compression Based Classification of Biological Sequences and Structures: Theory versus Practice. *BMC Bioinformatics*, 8:252, 2007.
- [33] P. Ferragina, A. Gulli. A personalized search engine based on web-snippet hierarchical clustering. *Software Practice & Experience*, 38(2): 189–225, 2008.
- [34] P. Ferragina, A. Isolani, D. Lombardini, T. Schiavinotto. Tauro: un sistema di ricerca e gestione avanzata di documenti XML. *Storicamente: Procs di un seminario sulla filologia digitale alla Columbia University*, Gedit edizioni, 4:2008 (ISSN: 1825-411X).
- [35] P. Ferragina, R. Gonzalez, G. Navarro, R. Venturini. Compressed Text Indexes: From Theory to Practice. *ACM journal on Experimental Algorithmics*, vol. 13, art. 12, February 2009.
- [36] P. Ferragina, R. Giancarlo, G. Manzini. The myriad virtues of wavelet trees. *Information and Computation*, 207(8): 849–866, 2009.

- [37] P. Ferragina, F. Luccio, G. Manzini, S. Muthukrishnan. Compressing and indexing labeled trees, with applications. *Journal of the ACM*, 57(1), 2009.
- [38] P. Ferragina, I. Nitto, R. Venturini. On Compact Representations of All-Pairs-Shortest-Path-Distance Matrices. *Theoretical Computer Science*, 411:3293–3300, 2010.
- [39] P. Ferragina, R. Venturini. The Compressed Permuterm Index. *ACM Transactions on Algorithms*, 7(1): article 10, 2010.
- [40] P. Ferragina, I. Nitto, R. Venturini. On optimally partitioning a text to improve its compression. *Algorithmica: Special issue on ESA09’s selected papers*, 61(1): 51-74, 2011.
- [41] P. Ferragina, U. Scaiella. Fast and accurate annotation of short texts with Wikipedia pages. *IEEE Software*, 29(1): 70-75, 2012.
- [42] P. Ferragina, T. Gagie, G. Manzini. Lightweight Data Indexing and Compression in External Memory. *Algorithmica: Special issue on LATIN10’s selected papers*, 63(3): 707-730, 2012.
- [43] P. Ferragina. On the weak-prefix search problem. *Theoretical Computer Science*, Special Issue on Selected Papers from the Conference CPM 11, Volume 483, 75-84, 2013.
- [44] P. Ferragina, J. Sirén, R. Venturini. Distribution-aware compressed full-text indexes. *Algorithmica: Special issue on ESA11’s selected papers*, 2013 (in corso di stampa).
- [45] P. Ferragina, I. Nitto, R. Venturini. On the bit-complexity of Lempel-Ziv compression. *SIAM Journal on Computing*, 42(4): 1521-1541, 2013.
- [46] P. Ferragina, R. Venturini. Compressed Cache-Oblivious String B-tree. *ACM Transactions on Algorithms*, 12(4), 2016.

(Chapters of) Books and Special Issues

- [47] P. Ferragina, F. Luccio. *Crittografia: principi, algoritmi, applicazioni*. Bollati Boringhieri, Torino, July 2001 (ISBN 88-339-5665-2).
- [48] P. Ferragina, R. Grossi, editors. *Proceedings of FUN ’04*. Editrice PLUS, 2004.
- [49] P. Ferragina. String search in external memory: algorithms and data structures. *Handbook of Computational Molecular Biology*, edited by Srinivas Aluru. Chapman & Hall/CRC Computer and Information Science Series, chapter 35, December 2005.
- [50] P. Ferragina, R. Grossi, F. Luccio, co-editors. *Special Issue on FUN ’04*. Theory of Computing Systems, 39(3), June 2006.
- [51] P. Ferragina. Il messaggio nella bottiglia: come scrivere i siti web per farsi trovare dai motori di ricerca. *Business Writing* di Alessandro Lucchini. Sperling & Kupfer, 2006.
- [52] F. Crestani, P. Ferragina, M. Sanderson, co-editors. *Proceedings of the Symposium on String Processing and Information Retrieval (SPIRE)*. Lecture Notes in Computer Science 4209, Springer Verlag, 2006.
- [53] P. Ferragina, G. Manzini, S. Muthukrishnan, co-editors. *Special Issue on the Burrows-Wheeler Transform*. Theoretical Computer Science, 387(3), 2007.
- [54] P. Ferragina. Editor of the area “*String Algorithms and Data Structures, Data Compression*”. Encyclopedia of Algorithms, Editor-in-Chief Ming-Yang Kao, Springer, 2008.
- [55] P. Ferragina is the (co)author of 5 entries of the *Encyclopedia of Algorithms*, Springer, 2008 [54], titled: *2D-pattern Indexing, Data Compression Boosting, Suffix Tree Construction in Hierarchical Memory, The Burrows-Wheeler Transform, Tree Compression and Indexing*.

- [56] F. Crestani, P. Ferragina, M. Sanderson, co-editors. *Information Retrieval journal: Special Issue on SPIRE '06*. Springer, (11)4, 2008.
- [57] P. Ferragina, G.M. Landau, co-editors. *Proceedings of the Symposium on Combinatorial Pattern Matching (CPM)*. Lecture Notes in Computer Science 5029, Springer Verlag, 2008.
- [58] P. Ferragina is the (co)author of 2 entries of the *Encyclopedia of Database Systems*, Springer Verlag, Editors-in-Chief Ling Liu and M. Tamer Özsu, titled: *Text compression* and *Indexing compressed texts*, 2009.
- [59] P. Ferragina, G.M. Landau, co-editors. *Theoretical Computer Science: Special Issue on CPM '08*. Springer, 410(51), November 2009.
- [60] P. Ferragina, F. Luccio. Ricercare su Internet. *Le Scienze: Algoritmi*. Eds G. Ausiello and R. Petreschi. Mondadori, 2011.
- [61] L. Epstein, P. Ferragina, co-editors. *Proceedings of the European Symposium on Algorithms (ESA)*. Lecture Notes in Computer Science 7501, Springer Verlag, 2012.
- [62] P. Ferragina, R. Venturini. Web Search. *The Power of Algorithms*. Eds G. Ausiello and R. Petreschi. Pages 107–137, Springer-Verlag, 2013 (ISBN 978-3-642-39651-9).
- [63] P. Ferragina. Capitolo su “I motori di ricerca”. *Documenti digitali*, Eds R. Guarasci and A. Folino. ITER, April 2013.
- [64] S. Leonardi, A. Panconesi, P. Ferragina, A. Gionis (Eds.). *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM)*. ACM Press, 2013.
- [65] L. Epstein, P. Ferragina, co-editors. *Algorithmica: Special Issue on Selected Papers of ESA 2012*. Algorithmica 70(3): 365-367 (2014).
- [66] A. Bernasconi, P. Ferragina, F. Luccio. *Elementi di Crittografia*. Pisa University Press, Pisa, 2015 (ISBN 978-88-6741-460-4).
- [67] P. Ferragina is (co)author of 5 entries of the *Encyclopedia of Algorithms*, Springer, 2016. Reviewed with respect to the ones edited in 2008 [55]. Titled: Burrows-Wheeler Transform (pp. 250-255), Compressing and indexing structured text (pp. 401-407), Indexed Two-Dimensional String Matching (pp. 973-977), Boosting textual compression (pp. 228-232), Suffix Tree Construction in Hierarchical Memory (pp. 2149-2154). (ISBN: 978-1-4939-2863-7)
- [68] P. Ferragina is (co)author of 2 entries of *Encyclopedia of Database Systems*, Springer, Editors-in-Chief Ling Liu and M. Tamer Özsu, published in 2009 [58] and reviewed in 2017 titled: *Text compression* and *Indexing compressed texts*.
- [69] P. Ferragina, S. Kurtz, S. Lonardi, G. Manzini. Chapter “Computational Biology”. *Handbook of Data Structures and Applications*, Eds D. Metha and S. Sahni. Springer, 2017 (ISBN 978-14-98-701853 / K24585).
- [70] P. Ferragina and F. Luccio. *Il Pensiero Computazionale: dagli algoritmi al coding*. Il Mulino, Bologna, 2017 (ISBN 978-88-15-27286-7).

Patents

- [71] The idea published in the paper [98] has been patented by **Lucent Technologies** (USA), with title “*Method and system for supporting multi-method dispatching in object-oriented programming*”, US Patent No. 6,434,566, 13 august 2002.

- [72] The idea published in the papers [116, 117] has been patented by Universities of Pisa and Rutgers, with title “*Method of structuring and compressing labeled trees of arbitrary degree and shape*”, US Patent No. 8,156,156, 10 April 2012.
- [73] The idea published in the paper [123] is under patenting by Yahoo!, 29 August 2007.
- [74] The idea published in the paper [125] is the subject of two US patents owned by Yahoo!, 30 September 2008. The one titled “*Query log mining for detecting spam hosts*” has been accepted (Patent no. 8,996,622, 31 March 2015).
- [75] US Provisional Patent Application on “Systems, methods and computer-accessible mediums for utilizing pattern matching in stringomes”, co-author B. Mishra (Courant Institute, New York University), owner New York University, 24 January 2013.

Proceedings of conferences

- [76] P. Bianchini, P. Ferragina, M. Notturmo Granieri, L. Tarricone. New techniques for speech understanding. *IEEE Conference on Acoustics, Speech and Signal Processing*, Minneapolis (USA), vol. 2, 127–130, 1993.
- [77] P. Ferragina, F. Luccio. Batch dynamic algorithms for two graph problems. *Parallel Architectures and Languages Europe (PARLE)*, Atene (Grecia), Lecture Notes in Computer Science 817, Springer-Verlag, 713–724, 1994.
- [78] P. Ferragina, A. Monti, A. Roncato. Trade-off between computational power and common knowledge in anonymous rings. *Colloquium on Structural Information and Communication Complexity*, Ottawa (Canada), 35–48, 1994.
- [79] P. Ferragina. Incremental Text Editing: a new data structure. *European Symposium on Algorithms (ESA)*, Utrecht (Olanda), Lecture Notes in Computer Science 855, Springer-Verlag, 495–507, 1994.
- [80] S. K. Das, P. Ferragina. An $o(n)$ work EREW parallel algorithm for updating MST. *European Symposium on Algorithms (ESA)*, Utrecht (Olanda), Lecture Notes in Computer Science 855, Springer-Verlag, 331–342, 1994.
- [81] P. Ferragina, R. Grossi. Fast Incremental Text Editing. *ACM-SIAM Symposium on Discrete Algorithms (SODA ‘95)*, San Francisco (USA), 531–540, 1995.
- [82] P. Ferragina. An EREW PRAM fully-dynamic algorithm for MST. *International Parallel Processing Symposium (IPPS ‘95)*, Santa Barbara (USA), 93–100, 1995.
- [83] P. Ferragina, R. Grossi. A fully-dynamic data structure for external substring search. *ACM Symposium on the Theory of Computing (STOC ‘95)*, Las Vegas (USA), 693–702, 1995.
- [84] P. Ferragina, R. Grossi. Optimal on-line search and sublinear time update in string matching. *IEEE Symposium on Foundations of Computer Science (FOCS ‘95)*, Milwaukee (USA), 604–612, 1995.
- [85] P. Ferragina, R. Grossi. Fast String Searching in Secondary Storage: Theoretical Developments and Experimental Results. *ACM-SIAM Symposium on Discrete Algorithms (SODA ‘96)*, Atlanta (USA), 373–382, 1996.
- [86] P. Ferragina. A simple parallel dictionary matching algorithm. *European Conference on Parallel Processing (EURO-PAR ‘96)*, Lione (Francia), Lecture Notes in Computer Science 1123, Springer-Verlag, 781–788, 1996.
- [87] P. Ferragina, F. Luccio. On the parallel dynamic dictionary matching problem: new results with applications. *European Symposium on Algorithms (ESA ‘96)*, Barcellona (Spagna), Lecture Notes in Computer Science 1136, Springer-Verlag, 261–275, 1996.

- [88] P. Ferragina, S. Muthukrishnan. Efficient dynamic method-lookup for object oriented languages. *European Symposium on Algorithms (ESA '96)*, Barcellona (Spagna), Lecture Notes in Computer Science 1136, Springer-Verlag, 107–120, 1996.
- [89] P. Ferragina, M. Montanero, R. Grossi. A note on updating suffix tree labels. *Italian Conference on Algorithms and Complexity (CIAC '97)*, Roma (Italia), Lecture Notes in Computer Science 1203, Springer-Verlag, 181–192, 1997.
- [90] L. Arge, P. Ferragina, R. Grossi, J. S. Vitter. On sorting strings in external memory. *ACM Symposium on the Theory of Computing (STOC '97)*, El Paso (USA), 540–548, 1997.
- [91] P. Ferragina, F. Luccio. Multi-string search in BSP. *Compression and Complexity of SEQUENCES 1997*, IEEE Press, Positano (Italia), 240–252, 1997.
- [92] L. Arge, P. Ferragina, R. Grossi, J. S. Vitter. Sequence sorting in secondary storage. *Compression and Complexity of SEQUENCES 1997*, IEEE Press, Positano (Italia), 331–346, 1997.
- [93] A. Czumaj, P. Ferragina, L. Gasieniec, S. Muthukrishnan, J. Träff. The architecture of a software library for string processing. *Workshop on Algorithm Engineering (WAE)*, Venezia (Italia), 166–176, 1997.
- [94] A. Crauser, P. Ferragina, K. Mehlhorn, U. Meyer, E. Ramos. Randomized external-memory algorithms for some geometric problems. *ACM Symposium on Computational Geometry*, Minneapolis (USA), 259–268, 1998.
- [95] M. Farach, P. Ferragina, S. Muthukrishnan. Overcoming the memory bottleneck in suffix tree construction. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Palo Alto (USA), 174–183, 1998.
- [96] A. Crauser, P. Ferragina. External memory construction of full-text indexes. *DIMACS Workshop on External Memory Algorithms and/or Visualization*, DIMACS (Rutgers University), 1998.
- [97] A. Crauser, P. Ferragina, K. Mehlhorn, U. Meyer, E. Ramos. I/O-optimal computation of segment intersections. *DIMACS Workshop on External Memory Algorithms and/or Visualization*, DIMACS (Rutgers University), 1998.
- [98] P. Ferragina, S. Muthukrishnan, M. deBerg. Multi-Method dispatching: A geometric approach with applications to string matching. *ACM Symposium on Theory of Computing (STOC)*, Atlanta (USA), 483–491, 1999.
- [99] S. Burkardt, A. Crauser, P. Ferragina, H.P. Lenhof, E. Rivals, M. Vingron. q -gram based database searching using a suffix array. *International Conference on Computational Molecular Biology (RECOMB)*, Lione (Francia), 77–83, 1999.
- [100] A. Crauser, P. Ferragina. On constructing suffix arrays in external memory. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 1643, Springer-Verlag, Praga (Repubblica Ceca), 224–235, 1999.
- [101] K. Brengel, A. Crauser, P. Ferragina, U. Meyer. An Experimental Study of Priority Queues in External Memory. *Workshop on Algorithmic Engineering (WAE)*, Londra (Inghilterra), Lecture Notes in Computer Science 1668, Springer-Verlag, 346–360, 1999.
- [102] R. Marangoni, A. Savona, P. Ferragina, N. Pisanti, L. Pagli, F. Luccio. *A method for paralogy trees construction*, German Conference on BioInformatics (GCB), Heidelberg, 2000.
- [103] P. Ferragina, G. Manzini. Opportunistic data structures with applications. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Redondo Beach (USA), 390–398, 2000.

- [104] P. Ferragina, G. Manzini. An experimental study of an opportunistic index. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Washington (USA), 269–278, 2001.
- [105] P. Ferragina, N. Koudas, S. Muthukrishnan, D. Srivastava. Two-dimensional substring indexing. *ACM Symposium on Principles of Database Systems (PODS)*, Santa Barbara (USA), 282–288, 2001.
- [106] G. Manzini, P. Ferragina. Engineering a lightweight suffix-array construction algorithm. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science vol. 2461, Springer-Verlag, Roma (Italia), 698–710, 2002.
- [107] V. Ciriani, P. Ferragina, F. Luccio, S. Muthukrishnan. Static optimality theorem for external memory string access. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Vancouver (Canada), 219–227, 2002.
- [108] P. Ferragina, G. Manzini. Compression boosting in optimal linear time using the Burrows-Wheeler Transform. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans (USA), 655–663, 2004.
- [109] M. Tavoni, E. Pierazzo, L. Leoncini, P. Ferragina, I. Boscaino, M. Tamosanis. An on-line Laboratory for Linguistic Research - Complete works of Dante lemmatized. *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH 2004)*, Goteborg (Svezia), 137–143, 2004.
- [110] P. Ferragina, A. Gulli. The Anatomy of SnakeT: a hierarchical clustering engine for web-page snippets. *European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Lecture Notes in Computer Science vol. 3202, Springer-Verlag, Pisa, 506–508, 2004.
- [111] P. Ferragina, A. Gulli. Experimenting SnakeT: a hierarchical clustering engine for web-page snippets. *European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Lecture Notes in Computer Science vol. 3202, Springer-Verlag, Pisa, 543–545, 2004.
- [112] P. Ferragina, G. Manzini, V. Mäkinen, G. Navarro. An alphabet friendly FM-index. *Symposium on String Processing and Information Retrieval (SPIRE)*, Lecture Notes in Computer Science vol. 3246, Springer-Verlag, Padova, 150–160, 2004.
- [113] P. Ferragina, A. Gulli. The Anatomy of a Hierarchical Clustering Engine for Web-page, News and Book Snippets. *IEEE Conference on Data Mining (ICDM)*, 395–398, Brighton (UK), 2004.
- [114] P. Ferragina, A. Gulli. A Personalized Search Engine Based on WebSnippet Hierarchical Clustering. *World Wide Web Conference (WWW)*, Tokio (Giappone), 801–810, 2005.
- [115] A. Farzan, P. Ferragina, G. Franceschini, J. Ian Munro. Cache-oblivious comparison-based algorithms on multisets. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science vol. 3669, Springer-Verlag, Eivissa (Spagna), 305–316, 2005.
- [116] P. Ferragina, F. Luccio, G. Manzini, S. Muthukrishnan. Structuring labeled trees for optimal succinctness, and beyond. *IEEE Symposium on Foundations of Computer Science (FOCS)*, Pittsburg (USA), 184–196, 2005.
- [117] P. Ferragina, F. Luccio, G. Manzini, S. Muthukrishnan. Compressing and searching XML data via two zips. *World Wide Web Conference (WWW)*, Edimburgh (UK), 751–760, 2006.
- [118] C. Corsi, P. Ferragina, R. Marangoni. The Bio-Prompt box. *ACM Intl. Conf. on Research in Computational Molecular Biology (RECOMB)*, Venezia, 2006 [poster].
- [119] P. Ferragina, R. Giancarlo, G. Manzini. The myriad virtues of wavelet trees. *International Colloquium on Automata, Languages and Programming (ICALP)*, Lecture Notes in Computer Science vol. 4051, Venezia, 561–572, 2006.

- [120] P. Ferragina, R. Giancarlo, G. Manzini. The Engineering of a Compression Boosting Library: Theory vs Practice in BWT compression. *European Symposium on Algorithms (ESA): Engineering and Applications Track*, Lecture Notes in Computer Science vol. 4168, Zurigo, 756–767, 2006.
- [121] P. Ferragina, R. Venturini. A simple storage scheme for strings achieving entropy bounds. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New Orleans (USA), 690–696, 2007.
- [122] P. Ferragina, J. Fischer. Suffix Arrays on Words. *Symposium on Combinatorial Pattern Matching (CPM)*, Lecture Notes in Computer Science vol. 4580, London (Ontario, CA), 328–339, 2007.
- [123] P. Ferragina, R. Venturini. Compressed Permuterm Index. *ACM SIGIR Conference*, Amsterdam (NL), 535–542, 2007.
- [124] P. Ferragina, R. Grossi, A. Gupta, R. Shah, J.S. Vitter. On Searching Compressed String Collections Cache-Obliviously. *ACM Principles of DataBase Systems (PODS)*, Vancouver (CA), 181–190, 2008.
- [125] C. Castillo, C. Corsi, D. Donato, P. Ferragina, A. Gionis. Query-log mining for detecting spam. *International Workshop on Adversarial Information Retrieval on the Web (AIR Web)*, Beijing (Cina), 2008.
- [126] A. Isolani, P. Ferragina, D. Lombardini, T. Schiavinotto. TauRo - A search and advanced management system for XML documents. *International Conference on Digital Humanities*, Oulu (Finlandia), 2008 (poster).
- [127] C. Castillo, C. Corsi, D. Donato, P. Ferragina, A. Gionis. Query-log mining for detecting polysemy and spam. *ACM Workshop on Web Mining and Web Usage Analysis (WebKDD)*, Las Vegas (USA), 2008.
- [128] P. Ferragina, I. Nitto, R. Venturini. On the bit-complexity of Lempel-Ziv compression. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, New York (USA), 768–777, 2009.
- [129] P. Ferragina, I. Nitto, R. Venturini. On optimally partitioning a text to improve its compression. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science vol. 5757, Copenhagen (DK), 420–431, 2009.
- [130] P. Ferragina, G. Manzini. On compressing the textual web. *ACM International Conference on Web Search and Data Mining (WSDM)*, New York (USA), 2010.
- [131] P. Ferragina, T. Gagie, G. Manzini. Lightweight data indexing and compression in external memory. *Latin American Theoretical Informatics Symposium (LATIN)*, Lecture Notes in Computer Science vol. 5757, 6034, Oaxaca (Mexico), 697–710, 2010.
- [132] A. Cisternino, M. Coppola, P. Ferragina, and D. Morelli. Information processing at work: On energy-aware algorithm design. *Workshop on Work in Progress in Green Computing*, IEEE Green Computing Conference, Chicago (USA), 2010.
- [133] P. Ferragina. Data structures: time, I/Os, entropy, joules! Invited paper at the *European Symposium on Algorithms (ESA)*, LNCS 6347 (part 2), Springer, 1-16, Liverpool (UK), 2010.
- [134] P. Ferragina and U. Scaiella. TAGME: on-the-fly annotation of short text fragments (by Wikipedia entities). Poster at *ACM Conference on Information and Knowledge Management (CIKM)*, Toronto (CA), 1625-1628, 2010.
- [135] P. Ferragina. On the weak-prefix search problem. *Symposium on Combinatorial Pattern Matching (CPM)*, Lecture Notes in Computer Science 6661, Springer, 261-272, Palermo, 2011.
- [136] P. Ferragina. Beyond the bag-of-words paradigm to enhance information retrieval applications. *International Conference on Similarity Search and Applications (SISAP)*, Invited Talk, ACM Press, 3-4, Lipari, 2011.

- [137] P. Ferragina, J. Siren and R. Venturini. Distribution-aware compressed full-text indexes. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 6942, Springer, 760-771, Saarbrücken (DE), 2011.
- [138] U. Scaiella, P. Ferragina, A. Marino, M. Ciaramita. Topical Clustering of Search Results. *ACM International Conference on Web Search and Data Mining (WSDM)*, Seattle (USA), 223-232, 2012.
- [139] D. Vitale, P. Ferragina, U. Scaiella. Classification of short texts by deploying topical annotations. *European Conference on Information Retrieval (ECIR)*, Lecture Notes in Computer Science vol. 7224, Barcellona (Spagna), 376-387, 2012.
- [140] M. Cornolti, P. Ferragina, M. Ciaramita. A framework for benchmarking entity-annotation systems. *International World Wide Web Conference (WWW)*, Rio de Janeiro (Brasile), 249-260, 2013.
- [141] P. Ferragina, R. Venturini. Compressed Cache-Oblivious String B-tree. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science vol. 8125, Sophia Antipolis (France), 469-480, 2013.
- [142] A. Farruggia, P. Ferragina, A. Frangioni, R. Venturini. Bicriteria data compression. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Portland (USA), 1582-1595, 2014.
- [143] A. Farruggia, P. Ferragina, R. Venturini. Bicriteria data compression: efficient and usable. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science vol. 8737, Springer-Verlag, Wrocław, Poland, 406-417, 2014.
- [144] M. Cornolti, P. Ferragina, M. Ciaramita, S. Rued, H. Schutze. The SMAPH system for query entity recognition and disambiguation. *ACM International Workshop on Entity Recognition and Disambiguation (ERD)*, within the Conference ACM SIGIR 2014, 25-30, 2014. (Winner of the Short-Track competition)
- [145] F. Piccinno, P. Ferragina. From TagMe to WAT: a new entity annotator. *ACM International Workshop on Entity Recognition and Disambiguation (ERD)*, within the Conference ACM SIGIR 2014, ACM SIGIR Forum, 55-62, 2014.
- [146] Ricardo Usbeck, Michael Rder, Paolo Ferragina *at al.* GERBIL - General Entity Annotator Benchmark. *International World Wide Web Conference (WWW)*, Firenze, 1133-1143, 2015.
- [147] Paolo Ferragina, Francesco Piccinno, Rossano Venturini. Compressed indexes for string-searching in labeled graphs. *International World Wide Web Conference (WWW)*, Firenze, 322-332, 2015.
- [148] P. Ferragina, F. Piccinno, R. Santoro. On analyzing hashtags in Twitter. *International AAAI Conference on Web and Social Media (ICWSM)*, Oxford (UK), 2015.
- [149] M. Cornolti, P. Ferragina, M. Ciaramita, S. Rued, H. Schutze. A piggyback system for joint entity mention detection and linking in web queries. *International World Wide Web Conference (WWW)*, Vancouver (Canada), 567-578, 2016.
- [150] M. Ponza, P. Ferragina, F. Piccinno. Document aboutness via sophisticated syntactic and semantic features. *International Conference on Natural Language & Information Systems (NLDB)*, LNCS 10260, pp. 441-453, Liegi (Belgio), 2017.
- [151] M. Ponza, P. Ferragina, S. Chakrabarti. Two-stage framework for computing entity relatedness in Wikipedia. *ACM Conference on Information and Knowledge Management (CIKM)*, pp. 1867-1876, Singapore, 2017.