Paolo Ferragina

Curriculum Vitae

January 2006

1 Personal

Associate Professor
Dipartimento di Informatica
University of Pisa, Italy

Address: L.go B. Pontecorvo, 3, 56127 Pisa, Italy.
Tel: +39-050-22127-64
Fax: +39-050-22127-26
Email: `ferragina@di.unipi.it`
Web: `http://www.di.unipi.it/~ferragin`

Born June 27, 1969 in Catanzaro (Italy). Married, one daughter born the 25th July 2003.

2 Brief Bibliography

Paolo Ferragina is currently Associate Professor of Computer Science at the University of Pisa, Italy. He got his Laurea degree (summa cum laude, 1992) and his PhD (1996) in Computer Science from the University of Pisa. He got his Post-doc from the Max-Planck Institut für Informatik (Saarbrücken, 1997-98). From 1998 to 2000, he has been Assistant Professor at the University of Pisa. From 2000, he is Associate Professor. In May 2005, he got the "Italian habilitation" to Full Professor.

He was Invited Scientist at various research centers and universities: IBM Research Center (Rome), University of North Texas (Denton, USA), AT&T Bell Laboratories (USA), Max-Planck-Institut für Informatik (Saarbrücken, Germany), AT&T Shannon Labs (USA), Munich Information Center for Protein Sequences (Germany), Theoretical BioInformatics Division (Heidelberg, Germany).

His research results received some awards and one US Patent:

- The "1995 Best Land Transportation Paper Award" from IEEE Vehicular Technology Society, for his research at IBM, journal paper #2 in Sect. 3.
- The "1997 EATCS Doctoral Dissertation Thesis Award" from the EATCS Italian Chapter, for his Phd Thesis on "Dynamic Data Structures for String Matching Problems".
- The "1997 Philip Morris Award on Science and Technology" for his research on "data structures for massive data".
- His PhD Thesis was ranked among the *four finalists* of the *1997 ACM Doctoral Dissertation Thesis Award*.
- The "2002 Research Capital award" from the University of Pisa, for his research on "compression and indexing technology".

- The US Patent No. 6,434,566, 13 August 2002, titled "Method and system for supporting multi-method dispatching in object-oriented programming", authored by P. Ferragina and S. Muthukrishnan and owned by Lucent Technologies (USA). The technical ideas underlying this patent have been then published in the *Procs of the ACM Symposium on Theory of Computing (STOC)*, 1999.

He has been invited speaker at various conferences and meetings, the most recent ones are:

- Plenary Speaker at *International Conference on Combinatorial Pattern Matching* (CPM), Instanbul (Turkey), 2004.
- Plenary Speaker at the *Symposium on String Processing and Information Retrieval (SPIRE)*, Buenos Aires (Argentina), 2005.

He served as (co)editor of special issues of journals and as author of (chapters in) books:

1. Co-author of an Italian book on Cryptography, published by Bollati-Boringhieri, 2001.
2. Co-editor of the *Special Issue on FUN '04* to be published in the journal *Theory of Computing Systems* (formerly "Mathematical Systems Theory"), 2006 (to appear).
3. Co-editor of the Special Issue on the *Burrows-Wheeler Transform and its Applications*, to be published in Theoretical Computer Science, Springer Verlag, 2006 (to appear).
4. P. Ferragina. String search in external memory: algorithms and data structures. Chapter 35 in the *Handbook of Computational Molecular Biology*, CRC Press, Editor Srinivas Aluru, December 2005.
5. Area Editor for the *Encyclopedia on Algorithms*, Editor in chief Ming-Yang Kao, Springer Press, 2006 (to appear).
6. P. Ferragina. String algorithms and data structures. LNCS Tutorial book on *"Algorithms for Massive Datasets"*, editor G. Brodal, Springer Verlag, 2006 (to appear).

He served as Program Committee member of several International Conferences: Cpm '98, Wae '01, EuroWeb '01, Spire '01, Icalp '02, Cpm '02, Latin '04, Fun '04, Esa '04, Spire '05, Cpm '06.
He also chaired some conferences and workshops:

- Co-chair of the *International Conference on FUN with Algorithms*, Isola d'Elba (Italy), 2004.
- Co-organizer of the *DIMACS Working Group on "The Burrows-Wheeler Transform: Ten years later"*, Rutgers University, 2004.
- Co-organizer of the *Meeting on "Space-conscious algorithms"*, Bertinoro (Italy), 2006.
- Co-chair of the *String Processing and Information Retrieval* (SPIRE), Glasgow (UK), 2006.

He serves as reviewer of the topmost international journals on Theoretical Computer Science, and is a member of the board of reviewers for the *Israeli Science Foundation*, in the area of Computer Science/Algorithmics.
He is currently the leader of various national and international projects on the design of Compression tools, Indexing algorithms and data structures, as well Search Engines. He is a member of the *European Network of Excellence NEMIS on "Text Mining"*, started in September 2002.

He is the author of more than 70 publications on the topmost international journals and conferences in Computer Science. We report below his ten topmost and most recent publications, to highlight his main discoveries, referring the reader to the following detailed list, limited to the last 10 years. For other information please have a look at his Home Page.

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): 236-280, 1999.
2. M. Farach, P. Ferragina, S. Muthukrishnan. On the sorting complexity of suffix tree construction. *Journal of the ACM*, 2000, 47(6): 987-1011, 2000.
3. P. Ferragina, G. Manzini. Indexing compressed texts. *Journal of the ACM*, 52(4): 552-581, 2005.
4. P. Ferragina, R. Giancarlo, G. Manzini, M. Sciortino. Boosting textual compression in optimal linear time. *Journal of the ACM*, 52(4): 688-713, 2005.
5. P. Ferragina, R. Grossi. Optimal on-line search and sublinear time update in string matching. *SIAM Journal on Computing*, 27(3): 713-736, 1998.
6. P. Ferragina, F. Luccio. Dynamic Dictionary Matching in External Memory. *Information and Computation*, 146(2): 85-99, 1998.
7. P. Ferragina, N. Koudas, S. Muthukrishnan, D. Srivastava. Two-dimensional substring indexing. *Journal of Computer and System Sciences: Special issue on selected papers from ACM PODS 2001*, 66(4): 763-774, 2003.
8. V. Ciriani, P. Ferragina, F. Luccio, S. Muthukrishnan. Static optimality theorem for external memory string access. *Proceedings of the IEEE Symposium on Foundations of Computer Science (FOCS)*, 219-227, 2002.
9. P. Ferragina, F. Luccio, G. Manzini, S. Muthukrishnan. Structuring labeled trees for optimal succinctness and beyond. *IEEE Foundations of Computer Science (FOCS)*, 2005.
10. P. Ferragina, A. Gullì. A personalized search engine based on web-snippet hierarchical clustering. *WWW*, 801-810, 2005.

He has been the coordinator of some committees on "international cooperation" (like Erasmus/Socrates) and "teaching planning and organization" within the Department of Computer Science of the University of Pisa and during his stay at the Max-Planck-Institut für Informatik. Among others duties, he is currently heading the Teaching Committee of the "Informatica Umanistica" (Humanities Computing) degree at the University of Pisa.

As far as his teaching experience is concerned, he taught many courses at undergraduate and graduate levels both in Italy and worldwide, on topics ranging from Data Structures to Cryptography, from Compression to Information Retrieval.
He was the advisor of two PhD Thesis:

- Andreas Crauser on *LEDA-SM: External memory algorithms and data structures in theory and practice*, Max-Planck-Institut für Informatik (Saarbrücken), . Currently Andreas is responsible for the office in Saabrücken of the `Algorithmic Solutions Software Gmbh` (i.e. the developers of the `LEDA Library`).
- Antonio Gullì on *On two WebIR boosting tools: Ranking and Clustering*, Dipartimento di Informatica (Pisa), 2005. Currently, Antonio is the Director of the Advanced Search Products at ASK Jeeves.

2.1 When Theory meets Practice!

Ferragina's research is mainly devoted to the design, analysis and experimentation of algorithms and data structures for storing, compressing, searching and mining information from large amounts of textual data, like Web repositories, XML file collections, genomic and textual databases. We recall below some of his theoretical findings that found interesting applications, and refer the reader to `http://roquefort.di.unipi.it/~ferrax/software.html` for further and updated information.

- The *String B-tree* is the first data structure providing I/O-optimal access to a dynamic set of arbitrary long strings on disk [ref 4 above]. This result got the *Philip Morris Award* and was cited at page 489 of the 3rd Volume of the famous Knuth's book *"The Art of Computer Programming"* (second edition, 1998) and in the section *"Reminiscences on Influential Papers"* of SIGMOD Record, September 2001. Interesting variations and improvements on this idea have been published in [refs 1,5,6,7,8 above].
- The *Divide-and-Conquer Technique* introduced in [ref 2 above] has been the first approach allowing the optimal construction of full-text indexing data structures in various models of computations. This result solved a long standing open problem, and it has been followed by many other experimental studies and achievements, the most recent one has been included in the `LEDA v4.5 Library` (Algorithmic Solutions Software GmbH) to implement the *bzip*-compressor. Ferragina is currently collaborating to the development of the `Leda` package: "Compression and Full-Text Indexing" (see `http://www.algorithmic-solutions.com/enpartner.htm`).
- The *FM-index* is the first data structure able to support efficient substring searches within an indexed text using roughly the same space required to store the text in compressed form [ref 3 above]. Technically, the FM-index retrieves the *occ* occurrences of a pattern $P[1,p]$ in $T[1,n]$ taking $O(p+occ\log^{\epsilon} n)$ time and uses at most $5nH_k(T)+o(n)$ bits of storage, where $H_k(T)$ is the k-th order empirical entropy of T. This space occupancy is the optimal $\Theta(n)$ bits in the worst case, and is the surprising $o(n)$ bits for compressible texts. The novelty of this approach resides in the exploitation of the relationship between the suffix array data structure and the Burrows-Wheeler compression algorithm. The FM-index received the *"2002 Research Capital Award"* from the University of Pisa and in the December 2003 the *Dr Dobb's Journal*, a reference for professional software developers, dedicated an article to this data structure. The FM-index had applications in other fields, other that Compression and Indexing: computational biology [J. Healy *et al.* 2003, Genome Research Journal; Sadakane *et al.* 2001, Genome Informatics] and Machine Translation Systems [R. Brown 2004, AMTA Conference].
- A boosting technique, in very informal terms, can be seen as a method that, when applied to a particular class of algorithms, yields improved algorithms in terms of one or more parameters characterizing their performance in the class. Despite the massive literature produced in the Data Compression field since the foundations laid by Shannon, no general boosting technique was known until the result in [ref 4 above]. Our *Compression Booster* takes a memoryless compression algorithm (like Huffman or Arithmetic) and turns it into an algorithm with a better compression performance guarantee without any time slowdown in its asymptotic performance. The technique is inherently combinatorial, does not assume any prior probabilistic model about the string source, and it does not deploy any training, parameter estimation and learning. [ref 4] also settles analytically some long standing open problems about the algorithmic structure of *bzip*-compressors. This result has received much attention [see the DIMACS Workshop on "BWT ten years later", August 2004].
- The US Patent mentioned in the previous section concerns with the efficient implementation of method lookups in object-oriented programming languages (OOPL). Research has identified *multi-methods* as a powerful feature to be added to OOPLs, and several experimental OOPLs have multi-methods. In the patent we show that the multi-method dispatching problem can be transformed to a geometric problem on a multi-dimensional integer grid, for which we then develop the first data structure that uses near-linear space and has sub-logarithmic query time. Previous solutions either used exponential space or required time linear in the number of candidate methods. The performance of the proposed solution almost matches that of the best known algorithm for mono-method dispatching.

  This result has stimulated substantial research both in the algorithmic field and, surpris-

ingly, in the field of IP Networking and packet classification/filtering. In fact our multi-method lookup problem became important in the IP setting with the name of "2d packet classification e filtering". Here, our algorithmic solution has been named as *FIS-tree*, and its importance is well recognized [Gupta-McKeown, "Algorithms for packet classification", IEEE Network, March 2001].

3 Detailed list of publications

International Journals

1. C. Calabrò, P. Ferragina, M. Notturno Granieri. Recognition of hand-written rotated digits by neural networks. *Machine Vision and Applications*, Springer Verlag, vol. 8, 351–358, 1995.

2. P. Comelli, P. Ferragina, M. Notturno Granieri, F. Stabile. Optical recognition of motor vehicle license plates. *IEEE Transactions on Vehicular Technology*, vol. 44(4), 790–799, 1995.

3. P. Ferragina. Static and dynamic parallel computation of connected components. *Information Processing Letters*, vol. 50(2), 63–68, 1994.

4. 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.

5. 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.

6. P. Ferragina. Dynamic Text Indexing under string updates. *Journal of Algorithms*, vol. 22(2), 296–328, 1997.

7. 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.

8. P. Ferragina, R. Grossi, M. Montangero. A note on updating suffix tree labels. *Theoretical Computer Science*, vol. 201(1–2), 249–262, 1998.

9. P. Ferragina, F. Luccio. Dynamic Dictionary Matching in External Memory. *Information and Computation*, vol. 146(2), 85–99, 1998.

10. 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.

11. 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.

12. P. Ferragina, R. Grossi. Improved Dynamic Text Indexing. *Journal of Algorithms*, vol. 31(2), pp. 291–319, 1999.

13. 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.

14. S. K. Das, P. Ferragina. An EREW PRAM algorithm for updating minimum spanning trees. *Parallel Processing Letters*, 9(1), pp. 111–122, 1999.

15. 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.

16. 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.

17. 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.

18. P. Ferragina, G. Manzini. An experimental study of a compressed index. *Information Sciences: special issue on Dictionary Based Compression*, vol. 135(1-2), pp. 13–28, 2001.

19. A. Crauser, P. Ferragina. A theoretical and experimental study on the construction of suffix arrays in external memory. *Algorithmica*, 32(1):1–35, 2002.

20. 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.

21. 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.

22. G. Manzini, P. Ferragina. Engineering a lightweight suffix array construction algorithm. *Algorithmica*, 40(1):33-50, 2004.

23. P. Ferragina, G. Manzini. Indexing compressed texts. *Journal of the ACM*, 52(4):552-581, 2005.

24. P. Ferragina, R. Giancarlo, G. Manzini, M. Sciortino. Compression boosting in optimal linear time. *Journal of the ACM*, 52(4):688-713, 2005.

25. P. Ferragina, V. Mäkinen, G. Manzini, G. Navarro. Compressed Representations of Sequences and Full-Text Indexes. *ACM Transactions on Algorithms*. 2006 (to appear).

Books, Chapters and Special Issues

1. P. Ferragina, F. Luccio. *Crittografia: principi, algoritmi, applicazioni*, Bollati Boringhieri, Torino, Luglio 2001 (ISBN 88-339-5665-2). [in italian]

2. P. Ferragina. String search in external memory: algorithms and data structures. *Handbook of Computational Molecular Biology*, CRC Press, Editor Srinivas Aluru, 2005.

3. P. Ferragina. String algorithms and data structures. *LNCS Tutorial book on "Algorithms for Massive Data Sets"*, Lecture Notes in Computer Science, Springer-Verlag, Editor Gerth Brodal, 2006 (to appear).

4. Co-editor (with R. Grossi and F. Luccio) of the *Special Issue on FUN '04* to be published in the journal *Theory of Computing Systems* (formerly "Mathematical Systems Theory"), 2005.

5. Co-editor (with G. Manzini and S. Muthukrishnan) of the *Volume on the Burrows-Wheeler Transform and its Applications*, to be published in Theoretical Computer Science, Springer Verlag, 2006.

6. Area editor of the *Encyclopedia on Algorithms*, Editor in chief Ming-Yang Kao, Springer, 2006 (to appear).

International Conferences

1. P. Ferragina, F. Luccio. Batch dynamic algorithms for two graph problems. *Parallel Architectures and Languages Europe (PARLE)*, Lecture Notes in Computer Science 817, Springer-Verlag, 713-724, 1994.

2. P. Ferragina, A. Monti, A. Roncato. Trade-off between computational power and common knowledge in anonymous rings. *Colloquium on Structural Information and Communication Complexity*, 35–48, 1994.

3. P. Ferragina. Incremental Text Editing: a new data structure. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 855, Springer-Verlag, 495–507, 1994.

4. S. K. Das, P. Ferragina. An $o(n)$ work EREW parallel algorithm for updating MST. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 855, Springer-Verlag, 331–342, 1994.

5. P. Ferragina, R. Grossi. Fast Incremental Text Editing. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 531–540, 1995.

6. P. Ferragina. An EREW PRAM fully-dynamic algorithm for MST. *International Parallel Processing Symposium (IPPS)*, 93–100, 1995.

7. P. Ferragina, R. Grossi. A fully-dynamic data structure for external substring search. *ACM Symposium on the Theory of Computing (STOC)*, 693–702, 1995.

8. P. Ferragina, R. Grossi. Optimal on-line search and sublinear time update in string matching. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 604–612, 1995.

9. P. Ferragina, R. Grossi. Fast String Searching in Secondary Storage: Theoretical Developments and Experimental Results. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 373–382, 1996.

10. P. Ferragina. A simple parallel dictionary matching algorithm. *European Conference on Parallel Processing (EURO-PAR)*, Lecture Notes in Computer Science 1123, Springer-Verlag, 781–788, 1996.

11. P. Ferragina, F. Luccio. On the parallel dynamic dictionary matching problem: new results with applications. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 1136, Springer-Verlag, 261–275, 1996.

12. P. Ferragina, S. Muthukrishnan. Efficient dynamic method-lookup for object oriented languages. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 1136, Springer-Verlag, 107–120, 1996.

13. P. Ferragina, M. Montangero, R. Grossi. A note on updating suffix tree labels. *Italian Conference on Algorithms and Complexity (CIAC)*, Lecture Notes in Computer Science 1203, Springer-Verlag, 181–192, 1997.

14. L. Arge, P. Ferragina, R. Grossi, J. S. Vitter. On sorting strings in external memory. *ACM Symposium on the Theory of Computing (STOC)*, 540–548, 1997.

15. P. Ferragina, F. Luccio. Multi-string search in BSP. *Compression and Complexity of SEQUENCES 1997*, IEEE Press, 240–252, 1997.

16. L. Arge, P. Ferragina, R. Grossi, J. S. Vitter. Sequence sorting in secondary storage. *Compression and Complexity of SEQUENCES*, IEEE Press, 331–346, 1997.

17. A. Czumaj, P. Ferragina, L. Gasieniec, S. Muthukrishnan, J. Träff. The architecture of a software library for string processing. *Workshop on Algorithm Engineering (WAE)*, 166-176, 1997.

18. A. Crauser, P. Ferragina, K. Mehlhorn, U. Meyer, E. Ramos. Randomized external-memory algorithms for some geometric problems. *ACM Symposium on Computational Geometry (SoCG)*, 259–268, 1998.

19. M. Farach, P. Ferragina, S. Muthukrishnan. Overcoming the memory bottleneck in suffix tree construction. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 174–183, 1998.

20. A. Crauser, P. Ferragina. External memory construction of full-text indexes. *DIMACS Workshop on External Memory Algorithms and/or Visualization*, DIMACS (Rutgers University), 1998.

21. 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.

22. P. Ferragina, S. Muthukrishnan, M. deBerg. Multi-Method dispatching: A geometric approach with applications to string matching. *ACM Symposium on Theory of Computing (STOC)*, 483–491, 1999.

23. 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)*, 77-83, 1999.

24. A. Crauser, P. Ferragina. On constructing suffix arrays in external memory. *European Symposium on Algorithms (ESA)*, Lecture Notes in Computer Science 1643, Springer-Verlag, 224–235, 1999.

25. K. Brengel, A. Crauser, P. Ferragina, U. Meyer. An Experimental Study of Priority Queues in External Memory. *Workshop on Algorithmic Engineering (WAE)*, Lecture Notes in Computer Science 1668, Springer-Verlag, 346–360, 1999.

26. R. Marangoni, A. Savona, P. Ferragina, N. Pisanti, L. Pagli, F. Luccio. *A method for paralogy trees construction*, German Conference on BioInformatics (GCB), 2000.

27. P. Ferragina, G. Manzini. Opportunistic data structures with applications. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 390–398, 2000.

28. P. Ferragina, G. Manzini. An experimental study of an opportunistic index. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 269–278, 2001.

29. P. Ferragina, N. Koudas, S. Muthukrishnan, D. Srivastava. Two-dimensional substring indexing. *ACM Symposium on Principles of Database Systems (PODS)*, 282–288, 2001.

30. 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, 698-710, 2002.

31. V. Ciriani, P. Ferragina, F. Luccio, S. Muthukrishnan. Static optimality theorem for external memory string access. *IEEE Symposium on Foundations of Computer Science (FOCS)*, 219-227, 2002.

32. P. Ferragina, G. Manzini. Compression boosting in optimal linear time using the Burrows-Wheeler Trasform. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 655-663, 2004.

33. M. Tavoni, E. Pierazzo, L. Leoncini, P. Ferragina, I. Boscaino, M. Tavosanis. The lemmatized Dante's works project. *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (ALLC/ACH)*, 2004.

34. P. Ferragina, A. Gullì. 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, Springer-Verlag, 2004.

35. P. Ferragina, A. Gullì. 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, Springer-Verlag, 2004.

36. 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, Springer-Verlag, 2004.

37. P. Ferragina, A. Gullì. The Anatomy of a Hierarchical Clustering Engine for Web-page, News and Book Snippets. *IEEE Conference on Data Mining (ICDM)*, 2004.

38. P. Ferragina, A. Gullì. The Anatomy of a Hierarchical Clustering Engine for Web-page, News and Book Snippets. *World Wide Web Conference (WWW)*, Tokio (Japan), 801–810, 2005.

39. A. Farzan, P. Ferragina, G. Franceschini, J. Ian Munro. Cache-oblivious comparison-based algorithms on multisets. *European Symposium on Algorithms (ESA)*, Eivissa (Spain), 2005.

40. 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), 2005.