

Mathieu RAFFINOT, Ph.D.

Professional address : LIAFA (www.liafa.jussieu.fr)
Université Paris Diderot - Paris 7, Case 7014
75205 Paris Cedex 13
France
Tel : +33 (0)1 44 27 28 42
E-Mail : raffinot@liafa.jussieu.fr
Personal data : Married, Born December 4, 1973. French.

1 *Position held*

Oct. 2007 – present : Full time CNRS researcher in algorithmics, LIAFA, France.
Full time CNRS researcher in algorithmics in Laboratoire Poncelet
Oct. 2005 – 2007 : (www.mccme.ru/lifr), 11 street Bolchoï Vlassievski, 119 002 Moscow,
Russia.
Full time CNRS researcher in computational biology.
Oct. 2000 – 2004 : Consultant in text algorithms and computational biology for the com-
pany Gene-IT (France) until Dec 2002.
Consultant in text algorithms for the company Systran (France) in
2004.

2 *Education*

1996-1999 : University of Marne-la-Vallée
Ph.D. in Computer Science, October 1999.
Advisor: Maxime Crochemore
1993-1996 : Bachelor and Master degrees at the Ecole Normale Supérieure of Lyon, France.

3 *Research interests*

Computational biology, pattern matching, indexing structures, graphs comparisons.

4 *Languages*

French : Native.
English : Fluent
German : Good basis (learned 10 years at school but never practiced since 1991).
Spanish : Basis (period of instruction of one month (4h/day) in August 98 in the Institut
Goethe of Santiago de Chile).
Russian : Basis.

5 *Visits of laboratories*

Year	Day and Month	Site	Collaborators
2007	2 to 21 Aug.	UQAM, Montréal, Canada	Anne Bergeron
2005	5 to 20 Nov.	UQAM, Montréal, Canada	Anne Bergeron
	30 May to 12 June	Laboratoire Poncelet, Moscou	M. A. Tsfasman
	25 Apr. to 14 May	UQAM, Montréal, Canada	Anne Bergeron
2004	21 Oct. to 8 Nov.	UQAM, Montréal, Canada	Anne Bergeron
2003	17 to 31 Aug.	University of Montréal, Canada	Nadia El-Mabrouk
	15 March to 17 Apr.	LIRMM, Montpellier, France	Christophe Paul Michel Habib
2002	3 to 22 Dec.	University of Montréal, Canada	Nadia El-Mabrouk
	10 to 27 Nov.	University of Chile, Santiago, Chile	Gonzalo Navarro
	20 Aug. to 13 Sep.	UQAM, Montréal, Canada	Anne Bergeron
	27 Apr. to 27 March	University of Montréal, Canada	Nadia El-Mabrouk
2001	3 to 8 Dec.	UMR 5558, Lyon	Marie-France Sagot Anne Bergeron
	5 to 10 Nov.	Institut Max-Planck, Berlin, Germany	Jens Stoye
	21 Apr. to 7 May	University of Montréal, Canada	Nadia El-Mabrouk
	4 Jan. to 7 March	University of Bar Ilan, Tel Aviv, Israel	Schmuel T. Klein
2000	20 Nov. to 20 Dec.	University of Chile, Santiago, Chile	Gonzalo Navarro
1998	24 July to 28 Sept.	University of Chile, Santiago, Chile	Gonzalo Navarro
1997	12 Nov. to 13 Dec.	University of Chile, Santiago, Chile	Gonzalo Navarro

6 *Conference (PC member)*

- 2008 Symposium on String Processing and Information Retrieval (SPIRE) (Australia, Oct.)
2007 Symposium on String Processing and Information Retrieval (SPIRE) (Santiago, Chile, Oct.)
Combinatorial Pattern Matching 2007 (London, Ontario, Canada, July)
2006 Symposium on String Processing and Information Retrieval (SPIRE) (Glasgow, U.K., Oct.)
2005 Combinatorial Pattern Matching 2005 (Seoul, Korea, June)

7 *Graduate student*

- 2002-2004 Co-adviser with Jean-Marc Champarnaud (LIFAR, Rouen, France) of Fabien Coulon: Ph.D. with honors the 12/06/2004.

8 *Publications*

8.1 *Book*

- [1] G. Navarro and M. Raffinot. Flexible pattern matching in strings : Practical On-Line Search Algorithms for Texts and Biological Sequences. *Cambridge University Press*, 2002, ISBN: 0-521-81307-7.

8.2 *Publications in journals*

- [2] A. Bergeron, C. Chauve, F. de Montgolfier, M. Raffinot. Computing common intervals of K permutations, with applications to modular decomposition of graphs. *SIAM Journal on Discrete Mathematics (SIDMA)*, to appear, 2008. (extended abstract in *European Symposium on Algorithms (ESA)*, LNCS 3669, Oct. 2005).

- [3] R. Kolpakov and M. Raffinot. New Algorithms for Text Fingerprinting. *Journal of Discrete Algorithms*, to appear, 2008. (extended abstract in CPM'2006, LNCS 4009).
- [4] F. Coulon and M. Raffinot. Fast Algorithms for Identifying Maximal Common Connected Sets of Interval Graphs. *Discrete Applied Mathematics*, 154(12):1709-1721, 2006.
- [5] S. Pasek, A. Bergeron, J.-L. Risler, A. Louis, E. Ollivier, and M. Raffinot. Identification of genomic features using microsynteny of domains: Domain teams. *Genome Research*, 15:867-874, 2005 (June).
- [6] G. Navarro and M. Raffinot. New Techniques for Regular Expression Searching. *Algorithmica* 41(2): 89-116, 2004.
- [7] M.-P. Béal, A. Bergeron, S. Corteel, and M. Raffinot. An algorithmic View of Gene Teams. *Theoretical Computer Science* 320(2-3):395-418, 2004. (partly appeared in WABI'2002, LNCS 2452, p. 464-476).
- [8] G. Navarro and M. Raffinot. A general practical approach to pattern matching over ziv-lempel compressed text. *Journal of Discrete Algorithms* 2(3):347-371, 2004. (extended abstract in CPM'99, LNCS 1645, p. 14-36).
- [9] N. El-Mabrouk and M. Raffinot. Approximate matching of secondary structures. *Journal of Bioinformatics and Computational Biology*. 2004. To appear. (extended abstract in RECOMB'2002, p. 156-164)
- [10] G. Navarro and M. Raffinot. Fast and Simple Character Classes and Bounded Gaps Pattern Matching, with Applications to Protein Searching. *Journal of Computational Biology* 10(6):903-923, 2003. (extended abstract in RECOMB'2001, Montréal, Canada, 2001)
- [11] N. Luc, J.-L. Risler, A. Bergeron, and M. Raffinot. Gene Teams: A New Formalization of Gene Clusters For Comparative Genomics. *Comput Biol Chem*, 27(1):59-67, Feb 2003.
- [12] M. Raffinot. A note on maximal repeats in strings. *Information Processing Letters*, 83, 2001.
- [13] G. Navarro and M. Raffinot. Fast and flexible string matching by combining bit-parallelism and suffix automata. *ACM Journal of Experimental Algorithmics (JEA)*, 5(4), 2000. <http://www.jea.acm.org>.
- [14] C. Allauzen and M. Raffinot. Simple optimal string matching. *Journal of Algorithms*, 36:102-116, 2000. (extended abstract in CPM'00, LNCS 1848, 2000)
- [15] M. Raffinot. Asymptotic estimation of the average number of terminal states in dawgs. *Discrete Applied Mathematics*, 92:193-203, 1999. (extended abstract in SWP'97, Valparaiso, Chile, 1997)

8.3 *Conference proceedings (with review)*

- [16] F. Coulon and M. Raffinot. Identification of Maximal Common Connected Sets of Interval Graphs and Tree Forests In CompBioNets, Recife, Brazil. 2004.
- [17] M. Habib, C. Paul and M. Raffinot. Common connected Components of Interval Graphs. In CPM'2004, number 3109 in Lecture Notes in Computer Science, 2004.
- [18] L. Dudoignon, E. Glemet, H. C. Heus, and M. Raffinot. *High Similarity Sequence Comparison in Clustering Large Sequence Databases*. In CSB'2002. IEEE Computer Society Press, 2002.

- [19] G. Navarro and M. Raffinot. Compact DFA representation for fast regular expression search. In *Proceedings of the 5th Workshop on Algorithm Engineering*, number 2141 in Lecture Notes in Computer Science, pages 1–12, 2001
- [20] C. Allauzen, M. Crochemore, and M. Raffinot. Efficient experimental string matching by weak factor recognition. In *Proceedings of the 12th Annual Symposium on Combinatorial Pattern Matching*, number 2089 in Lecture Notes in Computer Science, pages 51–72, 2001
- [21] C. Allauzen, M. Crochemore, and M. Raffinot. Factor oracle: a new structure for pattern matching. In Jan Pavelka, Gerard Tel, and Miroslav Bartosek, editors, *Proceedings of the 26th Seminar on Current Trends in Theory and Practice of Informatics (SOFSEM'99)*, number 1725 in Lecture Notes in Computer Science, pages 291–306, 1999.
- [22] G. Navarro and M. Raffinot. Fast regular expression search. In J. S. Vitter and C. D. Zaroliagis, editors, *Proceedings of the 3rd Workshop on Algorithm Engineering (WAE'99)*, number 1668 in Lecture Notes in Computer Science, pages 198–212, 1999. Journal version submitted.
- [23] G. Navarro and M. Raffinot. A bit-parallel approach to suffix automata: Fast extended string matching. In *Proc. CPM'98*, Lecture Notes in Computer Science 1448, pages 14–33. Springer-Verlag, 1998.
- [24] M. Raffinot. On the multi backward dawg matching algorithm (MultiBDM). In R. Baeza-Yates, editor, *Proceedings of the 4th South American Workshop on String Processing*, pages 149–165, Valparaiso, Chile, 1997. Carleton University Press.

8.4 *Technical reports (unpublished)*

- [25] A.-T. Gai, M. Habib, C. Paul and M. Raffinot. Identifying Common Connected Components of Graphs. Technical Report RR-LIRMM 03-016, LIRMM, University of Montpellier 2, July 2003. Submitted.
- [26] C. Allauzen and M. Raffinot. Oracle des facteurs d'un ensemble de mots. Technical report 99-11, Institut Gaspard Monge, University of Marne-la-Vallée, 1999.

9 Activities as consultant

Sep. 2004 – Nov. 2004: Consultant for the company Systran

SYSTRAN (<http://www.systran.fr>, <http://www.systranet.com>) is a company specialized in language translation. Its software belongs to most known of this field.

MAIN WORK

(this description remains very general for confidentiality)

- Algorithmic expertise of the source code of the translation machine. Improvement proposals.

April 2000 – Dec 2002 : Consultant for the company Gene-IT

GENE-IT (<http://www.gene-it.com>) is a bioinformatics company, resulting from the INRIA, which develops software LASSAP^[tm] and BIOFACET^[tm]. LASSAP^[tm] is a software for parallel large scale comparisons of genetic sequences. BIOFACET^[tm] is devoted to the clustering of very large EST (Expressed Sequence Tags) or cDNA databases.

MAIN WORK

(this description remains very general for confidentiality)

- Fast EST searches in large sequence databases.
- Matching large sets of EST.
- EST Clustering (Software BIOFACET^[tm]).

Dec. 1999 – Jan. 2000: Consultant for the company Systran

MAIN WORK

(this description remains very general for confidentiality)

- Minimization by *ad hoc* algorithms of automata based on large dictionaries.