

# Algorithmique sur le texte et application à la recherche dans les séquences biologiques

Mathieu Raffinot  
*CNRS - Laboratoire Génome et Informatique*  
*Université d'Evry*  
*Tour Evry 2*  
*523 place des terrasses de l'Agora*  
*91034 Evry*  
*FRANCE*

## Résumé

Cette présentation a pour but de décrire les différentes approches algorithmiques utilisées pour la recherche de mots dans les textes et dans les séquences biologiques. Nous présentons les algorithmes dans leur contexte historique afin de comprendre pourquoi ils ont été développés, et pourquoi et comment certains d'entre eux ont été (ou doivent être) optimisés. Les principaux algorithmes (pas exhaustif) présentés dans ce cadre général sont : McCreight (1976), Knuth-Morris-Pratt (1977), Smith-Waterman (1981), Landau-Vishkin (1989), BLAST (Altschul et al, 1990), Myers (1999), Amihoud et al.(2000), Crochemore et al. (2002).

## Références (articles) :

- D.E. Knuth, J.H. Morris, Jr, and V.R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(1) :323–350, 1977.
- E.M. McCreight. A space-economical suffix tree construction algorithm. *Journal of Algorithms*, 23(2) :262–272, 1976.
- S.F. Smith and M.S. Waterman. Comparison of biosequences. *Adv. Appl. Math.*, 2, 482-489 , 1981

G.M. Landau and U.Vishkin. Fast parallel and serial approximate string matching. *Journal of Algorithms*, 10(2) :157–169, 1989.

S.F. Altschul, W. Gish, W.Miller, E.W. Myers, and D.J. Lipman. BLAST : A basic local alignment search tool. *Journal of Molecular Biology*, 215 :403–410, 1990.

A. Amir, G. Landau, D. Keselman, M. Lewenstein, N. Lewenstein and M. Rodeh. Text Indexing and Dictionary Matching with One Error. *Journal of Algorithms*, 2000, Vol. 37, pp. 309–325.

M. Crochemore, G. M. Landau, and M. Ziv-Ukelson, A sub-quadratic sequence alignment algorithm for unrestricted cost matrices. *SODA'2002*, pages 679–688. ACM-SIAM, 2002.

**Références (livres) :**

D. Gusfield. *Algorithms on strings, trees and sequences : computer science and computational biology*. Cambridge University Press, Cambridge, 1997.

G. Navarro and M. Raffinot. *Flexible Pattern Matching in Strings — Practical on-line search algorithms for texts and biological sequences*. Cambridge University Press, 2002.

M. Crochemore and W. Rytter. *Jewels of Stringology*. World Scientific, Hong Kong, 2002, 310 pages.

M. Crochemore, C. Hancart et T. Lecroq. *Algorithmique du texte*. Vuibert, 2001 (en français)