

Algorithmics and pattern matching for biological purposes

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

Abstract

In this lecture we provide a general description of the main algorithmic approaches for searching patterns in texts and biological sequences. We place the algorithms in their historical contexts to highlight why they have been designed and why and how some of them have been (or should be) improved. The main algorithms we link together in this general framework are (not being exhaustive): McCreight (1976), Knuth-Morris-Pratt (1977), Smith-Waterman (1981), Landau-Vishkin (1989), BLAST (Altschul et al, 1990), Myers (1999), Amihod et al.(2000), Crochemore et al. (2002).

Some references (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 subquadratic sequence alignment algorithm for unrestricted cost matrices. *SODA '2002*, pages 679–688. ACM-SIAM, 2002.

Some references (books):

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 (in French)