

Reconnaissance de séquences dans une chaîne de caractères

2002

1 Introduction

Le problème de la détection de la présence d'une séquence dans une grande chaîne de caractères est un problème classique en informatique.

C'est une opération utile lors de l'édition de texte, mais aussi dans un ensemble d'autres domaines comme celui de la génétique, où il faut rechercher des séquences d'ADN (alphabet à 4 lettres), dans des chaînes de très grandes tailles.

2 Sujet

Le sujet consiste à implanter un algorithme général de recherche de séquences et à évaluer le temps moyen de détection d'une séquence dans une chaîne aléatoire.

Ce temps moyen dépend, entre autre, de l'algorithme, de la taille des chaînes mais aussi de la taille de l'alphabet, etc.

Une version parallèle de l'algorithme devrait permettre de diminuer ce temps moyen de recherche.

3 Indication

La détection de séquence est un problème classique abondamment traité dans la littérature. On pourra consulter la référence suivante ou faire une recherche sur le WEB :

- KNUTH, D.E., MORRIS JR, J.H., PRATT, V.R., 1977, Fast pattern matching in strings, *SIAM J.Comput.* 6(1):323-350