

Devoir n°1 : Bio-Algorithmique

Recherche exacte de motifs

Indications

Travail à effectuer **en binôme ou en monôme**
A envoyer au plus tard le **dimanche 16 avril 2023** à l'adresse recupspace@gmail.com
Le travail (compte rendu) doit être tapé dans un éditeur et enregistré au **format pdf**
Le code source dans un fichier à part
Nom de fichier = « Devoir1-nomsEtudiants.pdf »
L'objet du mail : « Devoir1-BioALGO »

L'objectif de ce devoir est d'implémenter les algorithmes de recherche de motifs dans un texte et d'analyser leur complexité théorique et expérimentale

1. Implémenter les algorithmes suivants pour rechercher toutes les occurrences de motifs dans un texte T de longueur n:
 - Algorithme de Boyer-Moore [BM] (recherche d'un seul motif M de longueur m)
 - Algorithme de Aho-Corasick [AC] pour la recherche multiple (un ensemble motifs S1, S2, ...Sk)
 - Algorithme de Rabin-Karp [RK] pour la recherche multiple (un ensemble motifs (S1, S2, ...Sk). Utiliser 3 fonctions de hachage (filtre de Bloom) de votre choix (voir TD).
2. Comparaison des 2 algorithmes (Aho-Corasick et Rabin-Karp)
 - Faites un ensemble de tests en incluant le **nombre de comparaisons effectuées** et le **temps d'exécution** des deux algorithmes.
Il faudra reporter les résultats dans des tableaux en tenant compte des paramètres suivants : *la taille du texte, les tailles des motifs (ou la taille globale des motifs)*.
Tracer les courbes correspondant (diagrammes) aux résultats collectés dans les tableaux.
 - Est ce que les résultats de test des algorithmes précédents sont en accord avec ce qui était attendu (complexité théorique)?
3. L'algorithme de Commentz-Walter [CW] est un algorithme de recherche multiple (plusieurs motifs) qui se base sur les idées de l'algorithme de Aho-Corasick et de l'algorithme de Boyer-Moore.
 - Décrire le principe de l'algorithme de Commentz-Walter
 - Illustrer son déroulement sur un exemple de recherche de motifs S1, S2, S3 dans un texte T
 - Effectuez plusieurs tests pour comparer les performances de l'algorithme [CW] et [AC]
 - Analysez les résultats de test (procéder comme dans la question 2).

Remettre **un rapport** contenant :

- le principe de chaque algorithme,
- les tableaux des tests,
- les courbes (ou diagrammes)
- une conclusion.
- **Le code source (bien commenté), dans un fichier à part**