

Devoir n°1 : Bio-Algorithmique

Recherche exacte de motifs

Indications

Travail à effectuer **en binôme ou en monôme**

A envoyer au plus tard le **jeudi 10 avril 2025** à l'adresse recupspace@gmail.com

Le travail (compte rendu) doit être tapé dans un éditeur et enregistré au **format pdf**

Nom de fichier = « Devoir1-nomsEtudiants.pdf »

L'objet du mail : « Devoir1-BioALGO »

L'objectif de ce devoir est d'implémenter les algorithmes de recherche **exacte** de motifs dans un texte et d'analyser leur complexité théorique et expérimentale sur des textes de différentes longueurs et des motifs différents.

1. Implémenter l'algorithme de Morris et Pratt [MP] (utilisant le tableau des bords B) pour la recherche de toutes les occurrences d'un **seul** motif M dans un texte T de longueur n, en prenant en compte le **nombre de comparaisons** effectuées. *PS : afficher le tableau des bords, le nombre de comparaisons et les positions des occurrences trouvées.*
2. Implémenter l'algorithme de Boyer Moore [BM] (utilisant le tableau dictionnaire) pour la recherche de toutes les occurrences d'un **seul** motif M dans un texte T de longueur n, en prenant en compte le **nombre de comparaisons** effectuées. *PS : afficher le tableau dictionnaire, le nombre de comparaisons et les positions des occurrences trouvées.*
3. Test et analyse des algorithmes [MP] et [BM]
 - Faire un ensemble de tests des deux algorithmes en reportant dans un tableau, les tailles du texte et du motif, le **nombre de comparaisons effectuées** et le **temps d'exécution pour différents exemples**. Montrer le *meilleur des cas* et le *pire des cas* et les cas qui donnent des résultats semblables.
4. Implémenter l'algorithme de Rabin-Karp [RK] pour la recherche **multiple** (un ensemble de motifs S1, S2, ...Sk, de même longueur), en utilisant les 2 variantes a. et b. suivantes :
 - a. Utiliser 1 seule fonction de hachage de votre choix (voir TD).
 - b. Utiliser 3 fonctions de hachage (filtre de Bloom) de votre choix (voir TD).

Test et analyse de l'algorithme [RK]

- Faire un ensemble de tests de l'algorithme [RK] en reportant le **nombre de comparaisons effectuées** et le **temps d'exécution pour différents exemples**, ainsi que le nombre de faux positifs.
- Analyser la performance de l'algorithme [RK], en termes du nombre de **faux positifs** induits en comparant les 2 variantes a/ et b/ (question précédente). Conclure.

5. Implémenter l'algorithme de Aho-Corasick [AC] pour la recherche multiple de toutes les occurrences d'un ensemble de motifs S1, S2, ...Sk. *PS : afficher l'automate des préfixes de motifs, la fonction de suppléance, la fonction de sortie et les positions des occurrences trouvées.*

Test et analyse de l'algorithme [AC]

- Faire un ensemble de tests de l'algorithme [AC] en reportant le **nombre de comparaisons effectuées** et le **temps d'exécution pour différents exemples** (en augmentant la taille du texte et le nombre de motifs)

PS : pour les différents tests des 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 la complexité théorique (voir cours) ?

6. L'algorithme de **Commentz-Walter** [CW] est un algorithme de recherche multiple (plusieurs motifs), il se base sur les idées des algorithmes de *Aho-Corasick* et de *Boyer-Moore*.

- Décrire le principe de l'algorithme de Commentz-Walter
- Illustrer son déroulement sur un exemple (ou plus) 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).

Rapport à remettre

Remettre **un rapport** avec la structure suivante :

- Une page de garde (en tête, titre , noms, ...)
- Une introduction dans laquelle on présente la bio-Informatique, ses applications et l'intérêt de la recherche exacte de motifs
- Le principe de chaque algorithme
- Les tableaux des tests,
- Les courbes (ou diagrammes)
- Une conclusion.
- **Une annexe** comportant le code source (bien commenté) des algorithmes