
Travail TP : La recherche du motif dans un texte.

1 Présentation du Problème

Soient :

- $T = t_1 t_2 \dots t_n$, une chaîne de caractères de longueur n sur un alphabet Σ , appelée le texte.
- $P = p_1 p_2 \dots p_m$, une chaîne de caractères de longueur m sur le même alphabet Σ , appelée le motif.

Objectif : Trouver tous les indices i tels que $1 \leq i \leq n - m + 1$ et $T[i : i + m - 1] = P$, c'est-à-dire :

$$\forall j \in [1, m], t_{i+j-1} = p_j.$$

Sortie : L'ensemble des indices $\mathcal{I} = \{i_1, i_2, \dots, i_k\}$, où chaque i_j est une position dans T telle que P apparaît exactement.

Objectifs :

- Implémenter plusieurs structures de données pour résoudre ce problème.
- Analyser et comparer les complexités en temps et en espace de chaque approche de manière empirique.

2 Questions

- **Développement d'Algorithmes**

1. Implémenter une solution basée sur la structure de tableau de suffixes. Analyser la complexité en temps et en espace de cette solution des différentes algorithmes en place (Construction, Tri et recherche).
2. Implémenter une solution en utilisant la structure de l'arbre de suffixes (Algorithme d'Ukkonen). Etudier la complexité pratique temporelle et spatiale de cette solution.
3. Implémenter la solution qui utilise le FM-index et étudier la complexité expérimentale temporelle et spatiale de cette solution (Etudier tous les algorithmes et toutes les structures de données.).

- **Analyse Empirique de la Complexité**

1. **Comparaison de la Performance :**

- Choisissez un ensemble de cas de test avec des longueurs de chaîne variables (Il faut aller au maximum).
- Enregistrez et analysez le temps d'exécution et l'utilisation de la mémoire de chaque approche pour chaque taille d'entrée.
- Créez des graphes pour montrer la croissance empirique des performances en temps et en espace en fonction de la taille de l'entrée.

2. **Exploration d'applications du monde réel :**

- Considérer un exemple à partir de données du monde réel (ADN, chaînes binaires, textes anglais, etc)
- Décrire comment la taille et la complexité des données du monde réel peuvent influencer les performances de chaque structure (à partir de l'étude faite sur les différents exemples).

3 Livrables

- **Code** : Remettre code de chaque solution et l'application.
- **Rapport** : Inclure un rapport (max 8 pages) qui couvre :
 - L'analyse de la complexité et les résultats empiriques (graphiques inclus). En spécifiant l'environnement d'expérimentation (Matériel et Logiciel).
 - La discussion sur les optimisations et les conditions qui réduisent la complexité pour l'application.

Ce travail doit être remis au plus tard Le 12/12/24 à 00:00.
Ce travail se fait en *trinôme ni plus ni moins*.