

Algorithmique et Programmation
TD n° 3 : Recherche de motifs

Exercice 1. Supposons que le motif P et le texte T sont des chaînes de caractères de longueurs respectives m et n dont les caractères sont tirés uniformément aléatoirement et indépendamment dans un alphabet à k éléments ($k \geq 2$). Montrer que le nombre espéré de comparaisons de caractères effectuées par l'algorithme naïf de recherche de motif est :

$$(n - m + 1) \frac{1 - k^{-m}}{1 - k^{-1}} \leq 2(n - m + 1).$$

Exercice 2. Soit Σ un alphabet. Supposons que chaque caractère $\sigma \in \Sigma$ a un code unique $c(\sigma) \in \mathbb{N}$ (par exemple, son code ASCII). Étant donnés deux entiers p et q fixés, nous définissons la fonction de hachage suivante sur les mots de Σ^* :

$$H(x_1 \dots x_n) = \left(\sum_{i=1}^n c(x_i) p^{n-i} \right) \bmod q.$$

1. Rappeler l'algorithme de Rabin-Karp qui cherche les occurrences d'un motif P dans un texte T en utilisant la fonction H . Calculez sa complexité dans le pire cas. Calculez sa complexité en moyenne, en supposant que H répartit équitablement tous les mots sur $\{0, \dots, q - 1\}$.
2. Soit S un ensemble de motifs. Modifier l'algorithme de Rabin-Karp pour chercher toutes les occurrences d'un motif quelconque de S dans un texte T .
3. Une *image* (un *motif*) est une matrice rectangulaire dont les éléments sont dans Σ . Soient T une image de taille $N \times M$ et soit P un motif de taille $n \times m$. Proposer un algorithme qui détermine si P est une sous-motif de T , c'est-à-dire, si il existe $(x, y) \in \mathbb{Z}^2$ tels que $T(x + i, y + j) = P(i, j)$ et $(x + i, y + j) \in \llbracket 1, N \rrbracket \times \llbracket 1, M \rrbracket$ pour tous les $(i, j) \in \llbracket 1, n \rrbracket \times \llbracket 1, m \rrbracket$.

Exercice 3. Donner la table de décalage pour la recherche des motifs suivants : *abcababcac* et *aabaaab*.

Exercice 4. La distance entre deux mots est le nombre minimum d'opérations nécessaires pour transformer un mot en l'autre, où une opération est une insertion, une suppression ou un échange entre lettres adjacentes. Proposer un algorithme qui calcule la distance entre deux mots.