NSI · TERMINALE / COURS

La recherche textuelle

Astrolab Cours · astrolabcours.fr

01 Le problème

Chercher un mot avec Ctrl+F dans un livre, ou une séquence de 20 lettres dans un génome de 3 milliards de bases. L'idée simple, essayer toutes les positions, devient trop lente sur de très longs textes.

01 L'idée : le pochoir qui glisse

On fait glisser un pochoir portant le motif le long de la page. Méthode naïve : on avance toujours d'une case. Méthode maline : on regarde la lettre de la page sous le **bout** du pochoir. Si cette lettre n'existe pas dans le motif, aucune position qui la recouvre ne peut marcher : on saute d'un coup toute la longueur du pochoir.

01 La recherche naïve

```
def recherche_naive(texte, motif):
    n, m = len(texte), len(motif)
    positions = []
    for i in range(n - m + 1):
        j = 0
        while j < m and texte[i + j] == motif[j]:
            j += 1
        if j == m:
            positions.append(i)
    return positions
```

Dans le pire des cas, environ $n \times m$ comparaisons.

Boyer-Moore-Horspool

1. On aligne le motif au début du texte et on compare **de droite à gauche**.
2. Si tout correspond, on a trouvé une occurrence.
3. Sinon, on regarde la lettre du texte alignée avec la **dernière** case du motif, et on décale le motif selon la **table des décalages**.

La table donne, pour chaque lettre, sa distance à la fin du motif (dernière case exclue). Une lettre absente du motif fait décaler de toute la longueur m.

Pour le motif `CGAT` :

© Astrolab Cours · astrolabcours.fr/nsi/terminale/recherche-textuelle/ · Tous droits réservés : usage personnel uniquement, republication et usage commercial interdits

Lettre	C	G	A	autre (dont T)
Décalage	3	2	1	4

Plus le motif est long et l'alphabet varié, plus les sauts sont grands : on lit souvent moins d'une lettre sur deux du texte.

Les pièges

- Comparer de gauche à droite dans Boyer-Moore : la règle des décalages ne marche plus.
- Dans la table, la dernière lettre du motif n'est pas prise en compte (sinon décalage nul).
- Une lettre absente du motif donne le décalage maximal m, pas 1.

© Astrolab Cours · astrolabcours.fr/nsi/terminale/recherche-textuelle/ · Tous droits réservés : usage personnel uniquement, republication et usage commercial interdits