

Algorithmique TD10: Recherche de motifs dans Chaînes de caractères

14 décembre 2009

1. Algorithme naïf

Supposons que le motif contienne une ou plusieurs occurrences d'un caractère joker pouvant remplacer un caractère arbitraire (caractère vide inclus).

- (a) Donner un algorithme polynômial en temps qui détermine si un tel motif apparaît dans un texte.

2. Algorithme de Rabin-Karp modulaire

Soit l'algorithme de Rabin-Karp non modulaire vu en cours suivant :

```
procedure rabin-karp (tableau P(M):entier , tableau T(N):entier ,
                    d:entier ,m:entier ,n:entier )

  variables
    h,m,t,i,s:entier
  h ←  $d^{m-1}$ 
  p ← 0
  t ← 0
  pour i ← 1 a tm pas 1
    p ←  $p + d^{m-i} * P(i)$ 
    t ←  $t + d^{m-i} * T(i)$ 
  fpour
  pour s ← 0 a n-m-1 pas 1
    si p=t alors
      ecrire("decalage valide: "& s)
    fsi
    t ←  $(t - T(s+1) * h) * d + T(s+m+1)$ 
  fpour
  si p=t alors
    ecrire("decalage valide: "& s)
  fsi
fprocedure
```

En pratique on ne peut pas utiliser cet algorithme pour des grandes valeurs de tm ou d . On va modifier l'algorithme ci-dessus afin de travailler modulo q en ne gardant que $p\%q$ et $t\%q$.

- (a) Modifier l'algorithme ci-dessus en conséquence. Attention aux décalages invalides (même valeur de p et t mais sous chaînes différentes).

- (b) Calculer sa complexité dans le pire des cas et illustrer un exemple pour lequel elle est atteinte.
- (c) Proposer un algorithme permettant de généraliser l'algorithme de Rabin-Karp modulaire afin de rechercher dans un texte une occurrence de n'importe lequel des k motifs d'un ensemble donné.