

Christiane Frougny

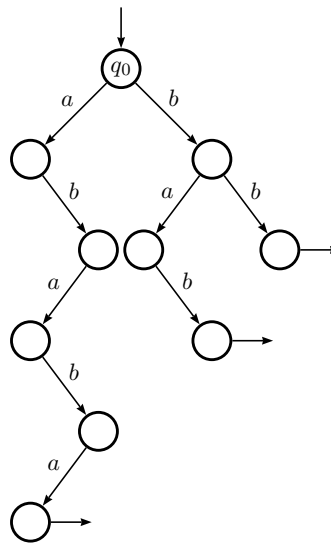
L3 : Algorithmique combinatoire
2^e semestre 2009–2010

Projet

**Recherche d'un ensemble fini de mots
par l'algorithme de Aho et Corasik**

Soit X un ensemble fini de mots sur un alphabet A , et t un texte sur A . On recherche si un ou plusieurs mots de X sont dans t .

Exemple $A = \{a, b\}$, $X = \{ababa, bab, bb\}$ et $t = aabababaaabb$. On définit l'*automate en arbre* de X ainsi



Soit $\mathcal{A} = (Q, A, q_0, T, \delta)$ l'automate (déterministe) en arbre de X , avec q_0 état initial, T ensemble des états terminaux, et δ la fonction de transition.

On dit qu'un mot u est un *préfixe* de X s'il existe un mot w de X tel que u soit préfixe de w .

On définit maintenant une fonction s pour un préfixe u de X par:

$s(u) = v$ où v est le plus long suffixe de u différent de u qui est aussi un préfixe de X .

Si q est l'état où on arrive après avoir lu un mot u dans l'automate en arbre, on associe u à q en notant $q = q_u$. On définit une *fonction de suppléance* s pour l'automate \mathcal{A} par $s(q_u) = q_v$ si et seulement si $s(u) = v$.

Le calcul de la fonction s s'obtient par l'algorithme suivant

Calcul des suppléments

début

```
soit  $L$  la liste des états  $q$  de  $Q$ ,  $q \neq q_0$ , ordonnée par niveau croissant de l'arbre;
 $s(q_0) := N$ ;
pour toute lettre  $\ell$  de  $A$  faire  $\delta(N, \ell) := q_0$ ;
pour  $q$  du premier au dernier élément de  $L$  faire
  {
     $z :=$  père de  $q$ ;
    soit  $\ell$  la lettre de  $A$  telle que  $\delta(z, \ell) = q$ ;
     $p := s(z)$ ;
    tant que  $\delta(p, \ell)$  non défini faire  $p := s(p)$ ;
     $s(q) := \delta(p, \ell)$ ;
    si  $\delta(p, \ell) \in T$  alors ajouter  $q$  à  $T$ ;
  }
```

fin

La recherche d'un mot de X dans t s'effectue ainsi. L'algorithme renvoie l'indice i dans t de la fin d'un mot de X , s'il y en a un, 0 sinon.

Algorithme de Aho et Corasik

début

```
 $q := q_0$ ;  $m := |t|$ ;
pour  $i$  de 1 à  $m$  faire
  {
    tant que  $q \neq q_0$  et  $\delta(q, t[i])$  non défini faire  $q := s(q)$ ;
    si  $\delta(q, t[i])$  non défini alors  $q := q_0$  sinon  $q := \delta(q, t[i])$ ;
    si  $q \in T$  alors Retourner( $i$ );
  }
```

Retourner(0)

fin

1) Implanter l'algorithme de Aho et Corasik pour trouver une occurrence d'un mot de X , et sa position dans le texte t .

2) Modifier l'algorithme pour trouver toutes les occurrences des mots de X dans t .

Rapport de projet

Remettre un rapport écrit contenant :

- une exposition du projet
- un listing commenté du programme
- un choix d'exécutions représentatives des différents cas.

Date de remise du projet : le jour de l'examen