Simulation d'un AFD à partir d'une expression régulière.

Leçons

Références : Dragon Book, Aho

Théorème

Soit E une expression rationnelle.
Après un prétraitement sur l'arbre syntaxique de E, on peut simuler l'AFD parcouru pour la lecture d'un mot w en temps $O(|w| \cdot |E|^2)$.

Remarque : une preuve montrant la possibilité de construire un AFD reconnaissant une expression rationnelle est de construire par récurrence un AFND la reconnaissant, puis de le déterminiser en construisant l'automate des parties.
La taille de l'automate obtenu est alors exponentielle, ce qui induit un algorithme de conversion s'exécutant en temps exponentiel.

Problème : cette borne est optimale, il suffit de regarder l'expression $(a|b)^* a (a|b)^{n-1}$. L'automate doit alors garder en mémoire les n dernières lettres, ce qui induit un nombre exponentiel d'états à garder en mémoire.
L'automate $A = (Q, q_0, F, \delta)$ est en effet minimal, avec :

$\Sigma = (a|b)^n$, $q_0 = b^n$, $F = a(a|b)^{n-1}$ et

$\delta(xw, y) = wy$ pour $w \in (a|b)^{n-1}$ et $x, y \in \{a, b\}$.

Et il a 2^n état !

Solution On peut éviter de générer l'automate complet en simulant uniquement le chemin emprunté, quitte à le garder en mémoire pour après.

Résumé

I - Construction d'un arbre syntaxique pour r#.

II - Calcul des fonctions suivantes :

eps (n) : l'expression enracinée en n reconnaît ε.

début (n) : feuilles ~~lettres~~ pouvant être au début d'un mot

fin (n) : feuilles ~~lettres~~ pouvant être à la fin.

suivant (n) : les positions pouvant correspondre à une lettre qui peut apparaître à la suite d'un mot reconnu par l'expression enracinée en n. l'ensemble des feuilles de l'arbre correspondant aux possibilités pour la lettre qui suit.

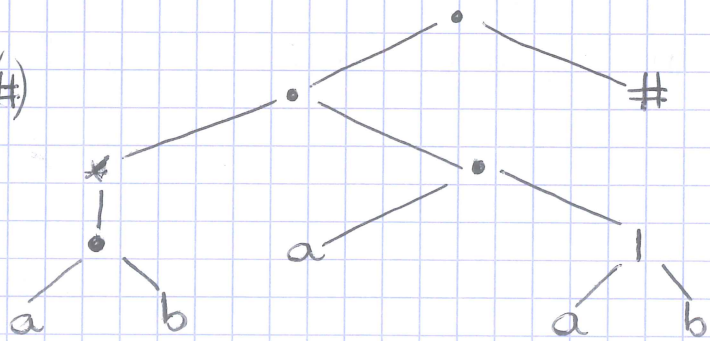
III - Construction, simulation et complexité de l'AFD.

I • Une expression rationnelle c'est :

* une lettre : a — une feuille
* $(r)^*$ — un noeud * avec un enfant r
* $(r_1 \mid r_2)$ — un noeud | avec deux enfants r_1 et r_2.
* $r_1 r_2$ — un noeud • avec deux enfants r_1 et r_2.

Cela définit bien un arbre, par induction structurelle sur les expressions rationnelles.

$((ab)^*(a(a \mid b))\#)$



On rajoute $\# \notin \Sigma$ à la fin pour savoir facilement en tête le moment où l'on arrive au bout d'un mot.

II | • On associe un numéro à chaque noeud de l'arbre généré, que l'on appellera position de ce noeud.

• Les fonctions eps, début, fin peuvent alors se calculer avec les règles suivantes :

Noeud n	eps(n)	début(n)	fin(n)
feuille ε	T	$\emptyset$	$\emptyset$
feuille $a \neq \varepsilon$ en position i	F	$\{i\}$	$\{i\}$
disjonction $r_1 \mid r_2$	$eps(r_1) \vee eps(r_2)$	$début(r_1) \cup début(r_2)$	$fin(r_1) \cup fin(r_2)$
concaténation $r_1 \cdot r_2$	$eps(r_1) \wedge eps(r_2)$	si $eps(r_1)$: $début(r_1) \cup début(r_2)$ sinon : $début(r_1)$	si $eps(r_2)$ $fin(r_1) \cup fin(r_2)$ sinon $fin(r_2)$
étoile r^*	T	début(r)	fin(r)

• La fonction suivant se construit alors facilement, car deux lettres consécutives d'un mot ne peuvent être obtenues que

- par concaténation : en avançant
- par une étoile : en reculant.

On obtient donc naturellement la fonction :

$$suivant(i) = \bigcup_{\substack{(\cdot, T_1, T_2) \\ i \in fin(T_1)}} début(T_2) \bigcup_{\substack{(*, T) \\ i \in début(T)}} fin(T)$$

III] • Il ne reste plus qu'à définir un AFD reconnaissant $r\#$ et une façon de calculer sa fonction de transition à la volée.

Cet automate est défini par :

* $Q = \mathcal{P}(\llbracket 1, n \rrbracket)$, où chaque nombre représente un des n nœuds de l'arbre syntaxique de $r\#$.
* $q_0 = \text{début}(n_0)$ où n_0 est la racine de T.
* $F = \{ q \in Q \mid n_\# \in q \}$ où $n_\#$ est la position de la feuille dont le symbole est $\#$.
* $\delta : Q \times \Sigma \longrightarrow Q$ tel que :

$$\delta(q, a) = \bigcup_{\substack{i \in q \\ \text{symbole}(i) = a}} \text{suivant}(i)$$

Remarque Cet automate, c'est vraiment un automate déterminisé, qui part de l'ensemble des états pouvant être initiaux, puis explore toutes les positions accessibles en lisant un mot.

Complexité : soit n la longueur de l'expression rationnelle r.

* prétraitement : calculs de eps, début, fin et suivant :
  $O(n)$ pour les trois premières
  $O(n^2)$ pour la dernière,

* parcours : $\delta(q, a)$ peut se calculer en $O(n^2)$ car $|q| = O(n)$ et $|\text{suivant}(i)| = O(n)$

$\Rightarrow$ on peut donc reconnaître $w \in \Sigma^*$ en $O(|w| \cdot n^2)$.