

Recherche d'un motif par un automate fini

Prérequis: Cours sur les langages formels

Motivation: Algorithme de recherche de motif avec un pré-traitement puis une recherche en temps linéaire dans le pire cas.

On pose: P le motif et T le texte, avec $|P| = m$ et $|T| = n$.

On souhaite construire un automate reconnaissant le motif P . On note: $P_n = P[1:n]$.

Définition: On appelle **fonction suffixe** associée à P et on note σ la fonction définie par:
$$\sigma(x) = \max \{h \in [0; m] \mid P_n \supseteq x\} \quad (\text{ou } \supseteq \text{ est la relation suffixe}).$$

Remarque: σ est bien défini car $P_0 = \varepsilon \supseteq x, \forall x \in \Sigma^*$.

Exemples: $P = ab$

• $\sigma(ccaca) = 1$

• $\sigma(ccab) = 2$

• $\sigma(abc) = 0$

Propriétés: $x, y \in \Sigma^*$.

• $\sigma(x) = m \Leftrightarrow P \ni x$

• $x \supseteq y \Rightarrow \sigma(x) \leq \sigma(y)$

Définition: On appelle **automate de recherche** associé à P l'automate fini défini par:

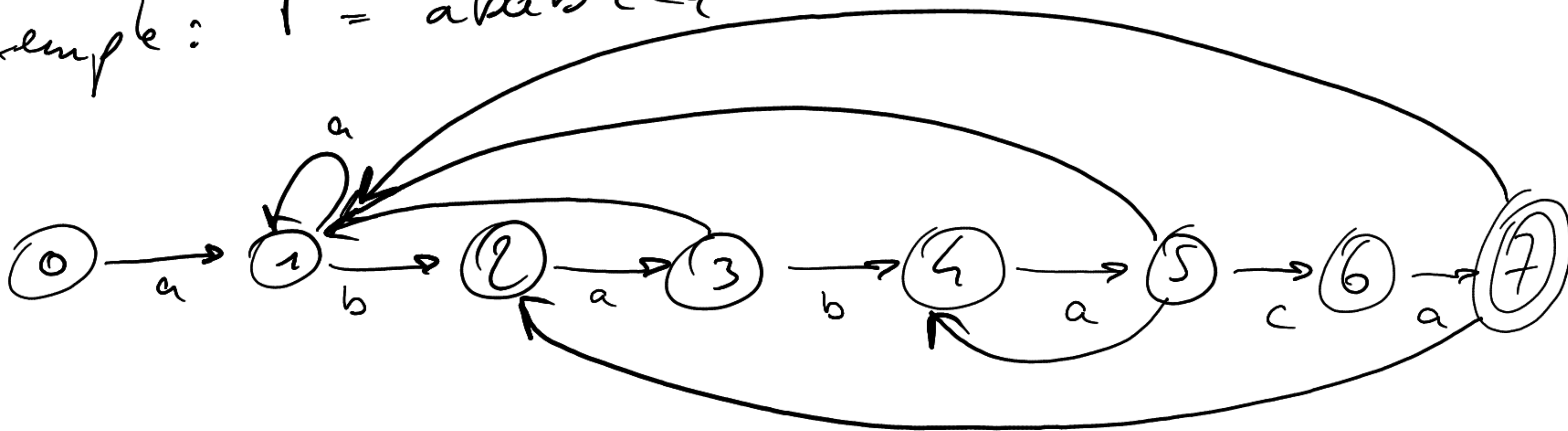
• $Q = \{0, 1, \dots, m\}$ états

• $q_0 = 0$ état initial

• $\delta(q, a) = \sigma(P_q a)$

Remarque: la fonction de transition est définie de cette manière car on veut pointer le plus long préfixe de motif P ayant jusqu'ici coïncidé avec la chaîne textuelle T .

Exemple: $P = ababaca$



(Toutes les autres transitions non représentées mènent vers $\emptyset$.)

i	-	1	2	3	4	5	6	7	8	9	10	11
$T[i]$	-	a	b	a	b	a	b	a	c	a	b	a
État	0	1	2	1	2	3	4	5	6	7	2	3

Fonction recherche - automate (T, δ, m) :

```

n = |T|
q = 0
Pour i de 1 à n
    q =  $\delta(q, T[i])$ 
    Si q = m
        Afficher "Le motif apparaît en position  $\{i-m\}^4$ "

```

Terminaison : OK

Complexité : $O(n)$ (sans calcul de δ)

Il reste à corriger.

Lemme: $\forall x \in \Sigma^*, \forall a \in \Sigma, \sigma(xa) \leq \sigma(x) + 1$.

Lemme: $\forall x \in \Sigma^*, \forall a \in \Sigma, \sigma(xa) = \sigma(p_{\sigma(x)} a)$.

Théorème: $\forall i \in [0; n], \delta^*(q_0, T_i) = \sigma(T_i)$.

(Pour le passage, recopiez les preuves sur du brouillon!)

Il ne reste plus que le calcul de δ :

Fonction calcul-fonction transition (P, Σ) :

```
m = |P|
pour q de 0 à m
  pour tout a ∈ Σ
    k = min(m+1, q+2)
    Répéter
      | k = k - 1
    jusqu'à  $P_k \supseteq P_q a$ 
     $\delta(q, a) = k$ 
  Retourner  $\delta$ 
```

Complexité : $O(m^3/|\Sigma|)$ (on peut améliorer ce précalcul à $O(m|\Sigma|)$).