

### TD N° 3 : Recherche de motifs

#### 1. Lecture gauche-droite : méthode de Morris & Pratt (AFD de $\Sigma^*x$ )

Soit le motif  $x = bababb$ .

**a.** Construire l'AFN reconnaissant  $\Sigma^*x$  à partir de l'écorché de  $x$  en ajoutant une boucle initiale sur toutes les lettres, puis l'AFD obtenu par déterminisation.

**b.** Simuler la recherche de  $x = bababb$  dans le texte  $t = babbabaabababbaababababab$  en indiquant les états d'arrivée de l'AFD lorsqu'on lit chaque lettre de  $t$ .

Les successions croissantes d'états consécutifs indiquent que la lecture s'effectue dans une fenêtre immobile. Les retours arrière dans l'AFD signalent un déplacement de la fenêtre vers l'avant.

En repérant les retours arrière dans la simulation ci-dessus, encadrer dans  $t$  les positions de la fenêtre de lecture de  $x$  testées au cours de la recherche.

**c.** Calcul de la fonction d'échec  $f$  de Morris & Pratt en appliquant pas à pas l'algorithme vu en cours :

$f(0) \leftarrow -1$

$i \leftarrow -1$

**pour**  $p = 1$  à  $n$

**tant que**  $i \geq 0$  et  $x(i+1) \neq x(p)$

$i \leftarrow f(i)$

$i \leftarrow i+1$

$f(p) \leftarrow i$

Grouper les valeurs dans un tableau.

**d.** Algorithme de Morris & Pratt : sans construction de l'AFD de  $\Sigma^*x$ . On lit  $t$  directement dans l'écorché de  $x$ . Quand dans un état donné  $p$  on ne peut pas lire la lettre courante du texte  $t$ , on réessaie à partir de  $f(p)$ , ou à partir de 0 si  $p = 0$ . Faire une simulation de cette méthode de lecture, en soulignant les passages par la fonction d'échec. Sous chaque état où elle est activée, indiquer le ou les états de remplacement proposés par  $f$ . Comparer avec la simulation de **b**.

#### 2. Lecture droite-gauche : AFD des suffixes de $x\sim$

On reprend le motif de l'exercice 1 :  $x = bababb$

**a.** Construire l'AFN pour les suffixes de  $x\sim = bbabab$  à partir de l'écorché de  $x\sim$  en rendant tous les états initiaux sauf le dernier, puis l'AFD obtenu par déterminisation.

**b.** Simuler la recherche de  $x = bababb$  dans le texte  $t = babbabaabababbaababababab$  en indiquant les états d'arrivée de l'AFD pour chaque lettre lue. Méthode : on essaie différentes positions de la fenêtre en commençant à gauche de  $t$ , mais on lit dans la fenêtre à partir de la droite en suivant un chemin dans l'AFD. En cas d'échec, on revient au dernier état final de l'AFD rencontré, et on se décale pour commencer une nouvelle lecture à partir du préfixe correspondant (suffixe de l'AFD), c'est-à-dire qu'on positionne le début de la fenêtre au niveau de cet état (et on lit de nouveau à partir de la fin de la fenêtre).

Encadrer dans  $t$  les positions de la fenêtre de lecture de  $x$  testées au cours de la recherche.

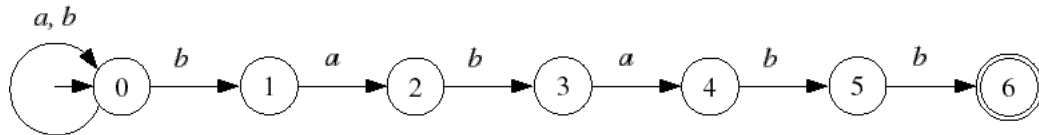
Comparer cette recherche avec celle effectuée à l'exercice 1.

### **3. Examen juin 2006**

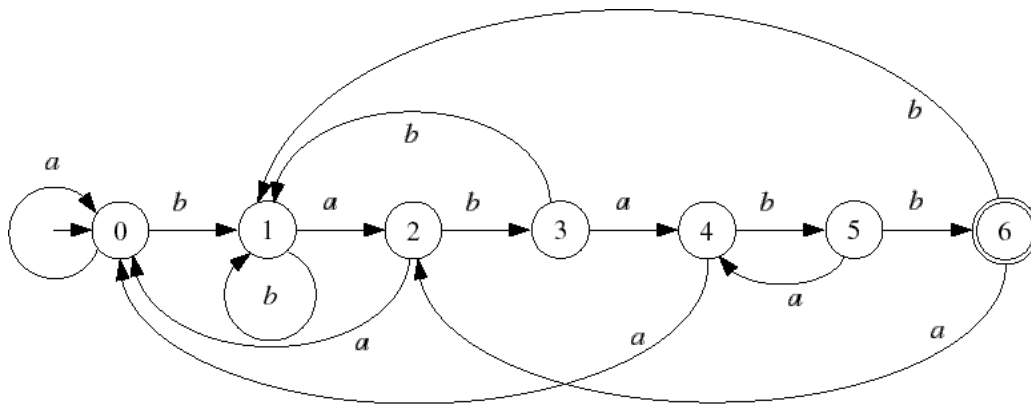
TD N° 3 - CORRECTION

1.

a. AFN pour  $\Sigma^*x$  obtenu à partir de l'écorché de  $x = bababb$



AFD après détermination (0 est inchangé) :  $1 = \{0, 1\}$ ,  $2 = \{0, 2\}$ ,  $3 = \{0, 1, 3\}$ ,  $4 = \{0, 2, 4\}$ ,  $5 = \{0, 1, 3, 5\}$ ,  $6 = \{0, 1, 6\}$



b. Lecture de  $t$  dans l'AFD de  $\Sigma^*x$

*b a b b a b a a b a b a b b a a b a b a b a b a b*  
 1 2 3 1 2 3 4 0 1 2 3 4 5 6 2 0 1 2 3 4 5 4 5 4 5

Le passage par l'état final 6 signale une occurrence du motif  $x = bababb$ .

Les suites croissantes d'états indiquent une fenêtre de lecture immobile. Un retour arrière indique un déplacement de la fenêtre vers l'avant : il y en a 6, d'où 7 positions testées.

1 2 3 [1 2 3 4 [0 1 2 3 4 5 6 [2 [0 1 2 3 4 5 [4 5 [4 5

<u>b a b b a b</u>	a a b a b a b b	a a b a b a b a b	a b a b a b
b a b <u>b a b a a b</u>	a b a b b a a b	a b a b a b a b a b	a b a b a b
b a b b a b a a	<u>b a b a b b</u>	a a b a b a b a b a b	a b a b a b
b a b b a b a a	b a b a b <u>b a a b a b</u>	a b a b a b	a b a b a b
b a b b a b a a	b a b a b b	a a <u>b a b a b a</u>	b a b
b a b b a b a a	b a b a b b	a a b a <u>b a b a b a</u>	b
b a b b a b a a	<u>b a b a b b</u>	a a b a b a	<u>b a b a b</u>

c. Calcul de la fonction d'échec : au départ  $f(0) = -1, i = -1$

$p = 1 \quad i = -1$ , on incrémente  $i = 0$ , et pose  $f(1) = 0$

$p = 2 \quad i = 0, x(1) = b \neq x(2) = a$ , on recule  $i = f(0) = -1$



