RICM1 – Automates et langages – devoir surveillé

Sans documents. Durée: 1h30.

12/12/2002

Exercice 1

1. Dans les cas ci-dessous, dire si le mot donné appartient au langage décrit par l'expression régulière :
   - 10100010; $(0^*10)^*$
   - 01110110; $(0+(11)^*)^*$
   - 000111100; $((011+11)^*(00)^*)^*$
2. Décrire en termes usuels les langages représentés par les expressions régulières $(1+10)^*$, $(0+10)^*(1+\epsilon)$.
3. Montrer, en fournissant une expression régulière correspondante, que les langages suivants sur $\Sigma = \{a,b,c\}$ sont réguliers :
   - tous les mots de longueur 2 ne contenant pas la lettre c;
   - tous les mots comportant un nombre impair d'occurrences de c;
   - tous les mots ne contenant pas ab comme facteur.

Indication. Justifiez vos réponses. En cas de difficultés vous pouvez toujours passer par les automates.

Exercice 2

Soit $\Sigma = \{a,b\}$, on considère deux langages suivants :
- L, le langage formé de tous les mots de Σ^* contenant aba;
- M, le langage défini par l'expression régulière $(b+aa^*bb)^*(\epsilon+aa^*+aa^*b)$.

1. Donner un automate non déterministe reconnaissant L. Déterminer l'automate minimal $\mathcal{A}$ reconnaissant L.
2. Donner un automate non déterministe avec ϵ-transitions reconnaissant M. Déterminer l'automate minimal $\mathcal{B}$ reconnaissant M.
3. En comparant les deux automates obtenus $\mathcal{A}$ et $\mathcal{B}$ déduire que $L = \overline{M}$.

Indication. N'oubliez pas que la minimisation s'applique aux automates déterministes uniquement.

Exercice 3

Montrer que le langage $\{a^nba^m | n < m\}$ n'est pas régulier.

Exercice 4

Soit $\mathcal{A}$ un automate avec k états, $L = L(\mathcal{A})$ – son langage. Montrer que L est infini si et seulement si il contient au moins un mot w de longueur supérieure à k.

Indication. "Seulement si" est facile. Pour prouver le "si" on peut utiliser le lemme de gonflage.

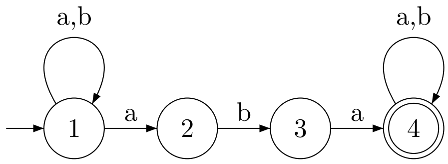
Corrigé

Exercice 1

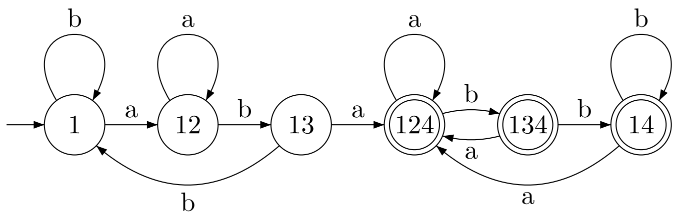
1. Dire si le mot donné appartient au langage décrit par l'expression régulière :
   - 10100010 : oui car (10)(10)((0)(0)(10))
   - 01110110 : non car tout bloc de 1 d'un mot de l'expression régulière $(0 + (11)^*)^*$ est de longueur paire.
   - 000111100 : oui car (00)((011)(11))(00)
2. Termes usuels des langages représentés par les expressions régulières suivantes :
   - $(1{+}10)^*$: tous les mots (binaires) sur 0,1, y compris ϵ, commençant par 1 et ne contenant pas 00.
   - $(0 + 10)^*(1 + \epsilon)$: tous les mots ne contenant pas deux 1 consécutifs.
3. Expressions régulières :
   - $(a + b)(a + b)$
   - Il y a deux solutions PcP ou bien $(a{+}b)^*cP$ où P est l'expression qui dénote un nombre pair de c, c'est-à-dire $P = (a + b + c(a + b)^*c)^*$
   - Après chaque bloc de a (sauf éventuellement un bloc à la fin) on doit avoir un c. D'où l'expression $(b + c + aa^*c)^*(\epsilon + aa^*)$, qui se simplifie à $(b + a^*c)^*a^*$.

Exercice 2

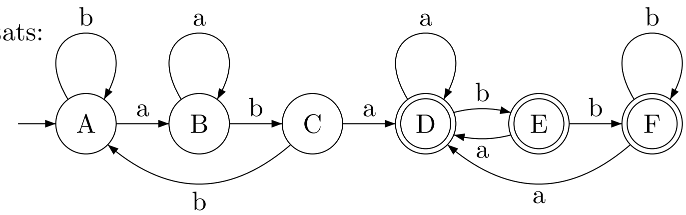
1. Un automate pour L:



On le déterminise:



On renomme les états:



Et on applique l'algorithme de minimisation. A l'initialisation on a

	B	C	D	E	F
A			X	X	X
B	-		X	X	X
C	-	-	X	X	X
D	-	-	-		
E	-	-	-	-	

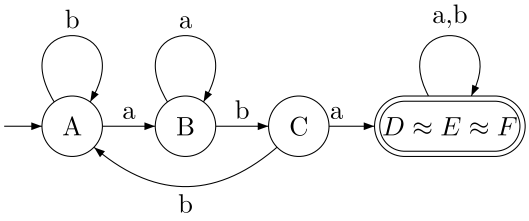
Ensuite on marque successivement

- la case (A,C) comme $A \xrightarrow{a} B$; $C \xrightarrow{a} D$ et la case (B,D) est déjà marquée;
- la case (B,C) comme $B \xrightarrow{a} B$; $C \xrightarrow{a} D$ et la case (B,D) est déjà marquée;
- la case (A,B) comme $A \xrightarrow{b} A$; $B \xrightarrow{b} C$ et la case (A,C) vient d'être marquée;

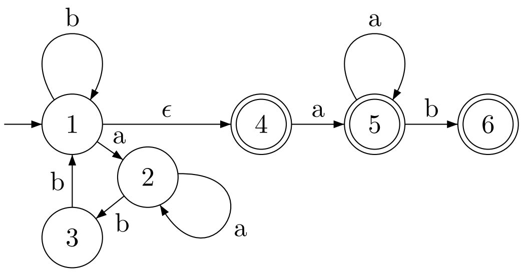
On vérifie facilement qu'il n'y a plus de cases à marquer et on obtient

	B	C	D	E	F
A	X	X	X	X	X
B	-	X	X	X	X
C	-	-	X	X	X
D	-	-	-		
E	-	-	-	-	

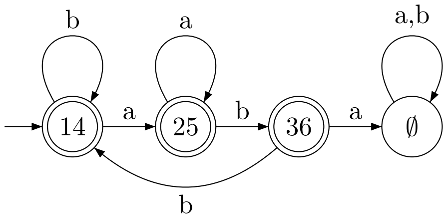
Cela donne quatre classes d'équivalence A, B, C et $D \approx E \approx F$. On construit l'automate minimal $\mathcal{A}$



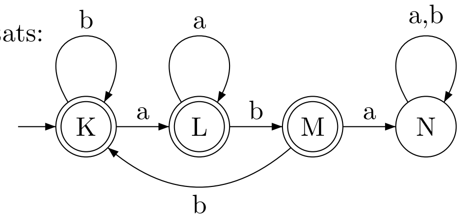
2. Un automate pour M:



On le déterminise:



On renomme les états:



Et on applique l'algorithme de minimisation. A l'initialisation on a

	L	M	N
K			X
L	-		X
M	-	-	X

Ensuite on marque successivement

- la case (K,M) comme $K \xrightarrow{a} L$; $M \xrightarrow{a} N$ et la case (L,N) est déjà marquée;
- la case (L,M) comme $L \xrightarrow{a} L$; $M \xrightarrow{a} N$ et la case (L,N) est déjà marquée;
- la case (K,L) comme $K \xrightarrow{b} K$; $L \xrightarrow{b} M$ et la case (K,M) vient d'être marquée;

On a marqué toutes les cases

	L	M	N
K	X	X	X
L	-	X	X
M	-	-	X

L'automate ci-dessus était déjà minimal.

3. On constate que la seule différence entre les automates déterministes $\mathcal{A}$ et $\mathcal{B}$ est que les états finals de l'un sont non-finals dans l'autre. D'où on peut déduire que leurs langages sont complémentaires: $L(\mathcal{A}) = \overline{L(\mathcal{B})}$, c'est-à-dire que $L = \overline{M}$.

Exercice 3

Pour démontrer que $L = \{a^m ba^n | m < n\}$ n'est pas régulier on utilise la méthode de preuve par contradiction.

Supposons que le langage L est régulier. Donc il existe un automate fini qui accepte L. Soit k le nombre d'états de cet automate.

On choisit un mot particulier $w = a^k ba^{k+1}$. Par définition du langage L on a $w \in L$.

Comme $|w| = 2k + 2 > k$, on peut appliquer le lemme de gonflement. Ce lemme dit, qu'il existe une décomposition $w = xyz$ avec $y \neq \epsilon$ et $|xy| \leq k$, telle que tous les mots de la forme xy^iz appartiennent aussi au langage L.

Comme $|xy| \leq k$, les morceaux x et u sont dans les k premiers caractères du mot w et ne peuvent contenir que des a. Soient r et $s \neq 0$ les nombres de lettres a dans x et y respectivement.

Donc, on a $x = a^r$, $y = a^s$, et, comme $w = xyz$, le dernier morceau z ne peut être autre chose que $a^{k-r-s}ba^{k+1}$.

Le mot "gonflé" $w' = xy^2z$ est de la forme xy^iz et **doit appartenir** au langage L. Mais

$$w' = xy^2z = a^r(a^s)^2 a^{k-r-s} ba^{k+1} = a^{r+2s+(k-r-s)} ba^{k+1} = a^{k+s} ba^k.$$

Or, comme $s \neq 0$, le nombre $k + s$ ne peut pas être inférieur à $k + 1$, et par définition du langage L, le mot $w' = a^{k+s}ba^k$ **ne peut pas appartenir** au même langage L. La contradiction obtenue conclue la preuve.

Exercice 4

Si. Supposons que L contient un w de longueur supérieure à k. En appliquant le lemme de gonflement on trouve qu'il existe une décomposition $w = xyz$ avec $y \neq \epsilon$ telle que pour tout $i \in \mathbb{N}$ on a $xy^iz \in L$.

Comme le mot y est non-vide, tous les mots xy^iz (correspondant aux différents i) sont différents. On a donc trouvé un nombre infini de mots différents dans L. Donc L est infini.

Seulement si. Supposons que L ne contient aucun mot de longueur supérieure à k. Donc tous les mots de langage L ont une longueur $\leq k$. Or il y a seulement un nombre fini de mots de telle longueur. On en déduit que L est fini.