Langages - Automates déterministes

2023 - Semestre 4 ; IUT Informatique Montpellier-Sète

Chapitre 1. Mots - Langages

1)Mots, langages

On se donne un **alphabet** fini;

par exemple X={a, b}

Alors:

- **un mot** sur *X* est ***une suite finie*** d'éléments de *X*. (*par ex. aaab, baa* ou *aaabb* qui peut se noter a^3b^2 …)
- il y a un **mot vide** noté ε.
- L'ensemble de tous les mots est noté *X**.

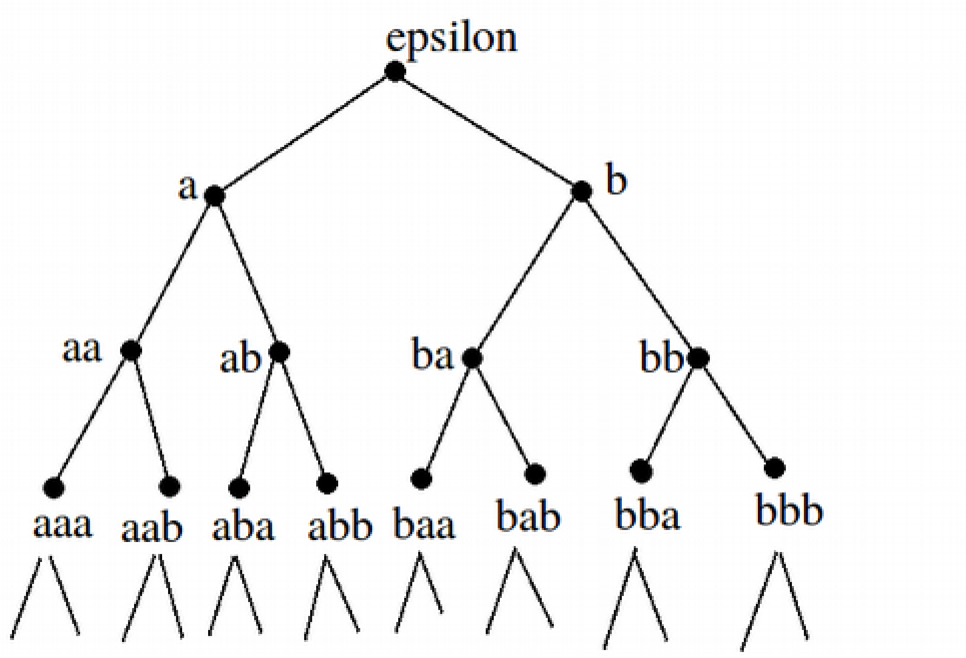
Notations:

- Soit m = abcba un mot sur l'alphabet X={a, b, c}.
  - **La longueur** de m est le nombre de « lettres » de m et est notée $|m|$(dans notre exemple$|m|$=5). On a $|\varepsilon|$= 0.
  - Pour i∈{1, 2, …, $|m|$}, on note m[i] la $i^{\text{ème}}$ lettre de m(dans notre exemple m[3]=c).
  - On note $|m|_a$ le nombre de a dans le mot m.

3

Représentation de X*

X={a,b}



Exemples de X*

1) X=Ø =>X*={ ε }

2)X={ | }=>X*={ ε,|, ||, |||,,||||,…}

3) X={ if, for, end, a, b, c, begin, ...}=>
X*={les programmes compilables ou non}

4) X={a,b,c,d,e,f,……,z}
=> X*={les mots ayant un sens ou non}

- On définit une loi interne sur X*:

la concaténation:

- Pour tout mot m de X* on a $m\bullet\varepsilon=\varepsilon\bullet m=m$
- Sinon, soient $m=m_1m_2\ldots.m_k$ et $n=n_1n_2\ldots.n_l$ deux mots de X* alors: $m\bullet n =m_1m_2\ldots.m_kn_1n_2\ldots.n_l$.

- Exemples:

  aabb •abb = aabbabb

  ab •ab •ab •ab = abababab=$(ab)^4$

6

Attention :

ε n'est pas une lettre.

L'écriture **aεbb** est incorrecte pour un mot.

Par contre on peut écrire **a•ε•bb=abb**

7

Le monoïde

*X** muni de l'opération de concaténation a une structure de **monoïde**

(i.e. • est ***associative*** et *X** possède un ***élément neutre*** pour cette opération).

Vocabulaire:

Soient **u**, **m** $\in$X*. **u** est un **facteur** de **m** lorsqu'il existe deux mots **v**, **w** $\in$ X* tels qu'on puisse écrire:

$$\mathbf{m = v \bullet u \bullet w}.$$

- Dans le cas particulier où **v = ε** (resp. **w = ε**), on dit que u est un **facteur gauche** (resp. **facteur droit**) ou encore un **préfixe** (resp. **suffixe**) de **m**.

Sous-mots

Soient u, m $\in A^*$.

u est un **sous-mot** de m lorsque u est une ***sous-suite extraite*** de m.

(donc m est un sur-mot de u)

Exemples :

On considère le mot **MANGER** :

- **ANG** est un facteur qui n'est ni préfixe ni suffixe
- **MA** est un facteur préfixe
- **GER** est un facteur suffixe
- **MAGR** est un sous-mot qui n'est pas un facteur.

MANGER est un facteur et un sous-mot.

De même **ε** est un facteur et un sous-mot.

Exercice :

On considère le mot « **manger** »

1) Combien a-t-il de préfixe ?

2) Combien a-t-il de facteurs ?

3) Combien a-t-il de sous-mots ?

12

factorisation

Soient m, f_1, f_2, ….f_k $\in A^*$ ($k \geq 1$).

La suite (f_1, f_2, ….f_k) est une **factorisation** de m lorsque m = $f_1 \bullet f_2 \bullet \ldots \bullet f_k$.

Exemple :

manger= ma•ng•ε•er donc

(ma, ng, ε, er) est une factorisation du mot manger en 4 facteurs.

15

- Un **langage** est un sous-ensemble de X*
-

par exemple: Soit X={a, b}:

L={ m $\in$X*/ $|m|_a$ = 2} ={ aa, aab, aba, baa,…}

- Remarques:
  - L'alphabet X est fini (donc ***dénombrable***)
  - L'ensemble X* de tous les mots est infini (sauf si X=∅) mais ***dénombrable*** (i.e. en bijection avec IN). Donc un langage est ***dénombrable***.
  - L'ensemble de tous les langages ***n'est pas dénombrable.***

16

Ordre lexicographique sur X*

On suppose que X est muni d'un ordre ***total*** $\leq_X$ (appelé ***ordre alphabétique***).

Ordre lexicographique sur X* ($\leq_{X^*}$):

1) Si m_1 est ***préfixe*** de m_2 alors $m_1 \leq_{X^*} m_2$.
2) Si $m_1 = u\bullet m'_1$ et $m_2 = u\bullet m'_2$ avec m'_1 et m'_2 $\neq \varepsilon$, et $m'_1[1] \neq m'_2[1]$, (éventuellement $u=\varepsilon$) alors: $\mathbf{m_1 \leq_{X^*} m_2}$ **ssi** $\mathbf{m'_1[1] \leq_X m'_2[1]}$,

Remarque: $\leq_{X^*}$ est total

17

Exemple :ordre lexicographique

X={a,b} avec a $\leq_X$ b :

ab $\leq_{X^*}$ abb

ababb $\leq_{X^*}$ abb

aaaaa $\leq_{X^*}$ b

aaab $\leq_{X^*}$ aab

18

Ordre hiérarchique sur X*

On classe les mots par longueur croissante

(un mot plus court est classé avant un mot plus long)

A longueur égale, les mots sont classés suivant l'ordre lexicographique.

Exemple :ordre hiérarchique

b $\leq_H$ aa

aab $\leq_H$ abb

aa $\leq_H$ aaa

aab $\leq_H$ aaab

abb $\leq_H$ abba

Exercice :

1) Donnez, dans l'ordre hiérarchique, les 5 premiers mots du langage

$$\mathbf{L=\{ m \in X^*/ |m|_a = 2\}}$$

2) Même question avec l'ordre lexicographique .

3) Reprendre les questions précédentes avec le langage

$$\mathbf{L'=\{ m \in X^*/ bb \text{ facteur de } m\}}$$

21

TD sur les langages formels

Exercice 1:

Soit $X = \{a, b, c\}$ un alphabet et $m = abca$ un mot sur X . Donner tous les facteurs, les suffixes, les préfixes et les sous-mots de m .

Exercice 2:

Soit $m = aba$. Combien y a-t-il de factorisations de m :

1. en deux facteurs différents de ε ?
2. en deux facteurs quelconques ?
3. en trois facteurs différents de ε ?
4. en trois facteurs quelconques ?

Exercice 3:

Soit $X = \{a, b\}$ un alphabet et C le sous-ensemble de mots de X^* ayant aba comme facteur et b^3 comme sous-mot.

1. Quelle est la longueur minimale k_0 des mots de C?
2. Combien y a-t-il dans C de mots de longueur k_0?
3. Soit $k \geq 0$ un entier. Quel est en fonction de k le nombre de mots de C de longueur inférieure ou égale à k et n'ayant pas a^3 comme sous-mot ?

Exercice 4:

Soit $X = \{a, b\}$ un alphabet muni de la relation d'ordre totale $\leq$ telle que $a \leq b$. On énumère les mots de X^* suivant l'ordre hiérarchique: le numéro 1 est le plus petit mot suivant l'ordre hiérarchique (donc ε), le numéro 2 est le suivant, ...

1. Donner les 10 premiers mots.
2. Soit k un entier strictement positif. Donner en fonction de k le nombre de mots de longueur strictement inférieure à k. En déduire le numéro de a^k.
3. Quel est le numéro de $m = abaabbbab$?
4. Quel est le mot de X^*dont le numéro est 300 ?

Exercice 5:

Soient X un alphabet, A et B deux langages sur X^*. On note $A \bullet B$ le langage de X^* défini par

$$A \bullet B = \{u \bullet v, u \in A \wedge v \in B\}$$

Pour les questions suivantes on prendra $X = \{a, b\}$. Calculer $A \bullet B$:

1. $A = \{a, ab, bb\}$ et $B = \{\varepsilon, b, aa\}$.
2. $A = \emptyset$ et$B = \{a, ba, bb\}$.
3. $A = \{\varepsilon\}$ et $B = \{b, aba\}$
4. $A = \{aa, ab, ba\}$ et $B = X^*$.

Exercice 6:

Soient $X = \{a, b\}$ un alphabet et L un langage sur X^*. On définit $Pref\,(L) = \{u \in X^*, \exists v \in L, u \,\text{préfixe de v}\}$. Déterminer $Pref\,(L)$ dans chacun des cas suivants:

1. $L_1 = \{a^n b^n, n \geq 0\}$.
2. $L_2 = \{a^n b^m, 0 \leq n \leq m\}$.
3. $L_3 = \{a^n b^m, 0 \leq m \leq n\}$
4. $L_4 = \{u \in X^*, |u|_a = |u|_b\}$

Chapitre 2. Automates déterministes

2) Automates déterministes

- **2a) <u>définitions</u>:**

  Un **automate fini** est un quintuplet

  $A = (X, E, e_0, F, \sigma)$ où:

  - X est un ensemble fini, appelé alphabet.
  - E est un ensemble fini, appelé ensemble des états. (X et E disjoints)
  - e_0 est un élément distingué de E, appelé état initial.
  - F est un sous-ensemble de E, appelé ensemble des états finaux.

- σ est une application de ExX dans E, appelée fonction de transition.

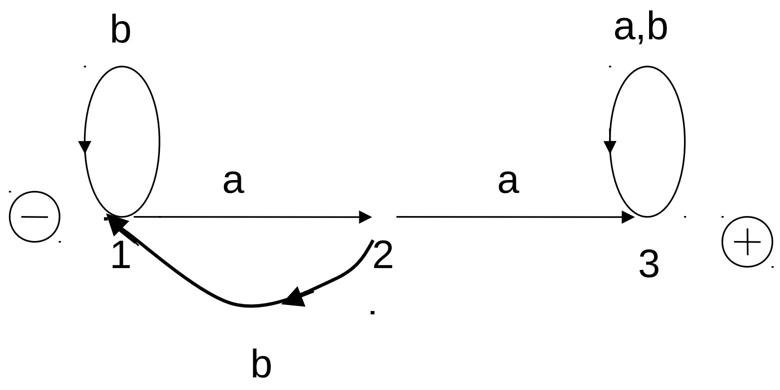
Exemple d'automate fini: A_1

$$X=\{a,b\};E=\{1,2,3\};e_0=1;F=\{3\};$$
$$\sigma:(1,a)\to 2;(2,a)\to 3;(3,a)\to 3;(1,b)\to 1;(2,b)\to 1;(3,b)\to 3;$$

	a	b
1	2	1
2	3	1
3	3	3

Table de transition

Graphe de transition



3

- <u>Définition 2</u>:

On étend la fonction de transition **σ** à une *application* **σ*** de ExX* dans E en posant, pour tout m ∈ X* et tout e ∈E:

- Si m=ε alors σ*(e,m)=e

-Si m = x • m' avec x∈X et m'∈X*, et si σ(e,x)=e',alors:

σ*(e,m)= σ*(e,x • m')= σ*(σ(e,x),m')= **σ*(e',m')**

Il est commode de noter e **x** m = σ*(e,m).

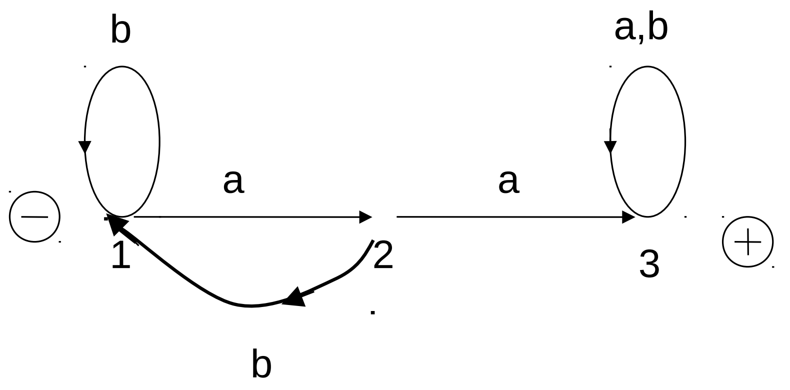
La dernière ligne de la définition s'écrit donc:
e **x** m=e **x** (x • m')= (e **x** x)**x**(m')=e' **x** m'

4

Exemple avec A_1

$\sigma^*(1,abaab)=\sigma^*(2,baab)=\sigma^*(1,aab)$
$=\sigma^*(2,ab)=\sigma^*(3,b)=\sigma^*(3,\varepsilon)=1$

On écrira: 1 x abaab = 3



5

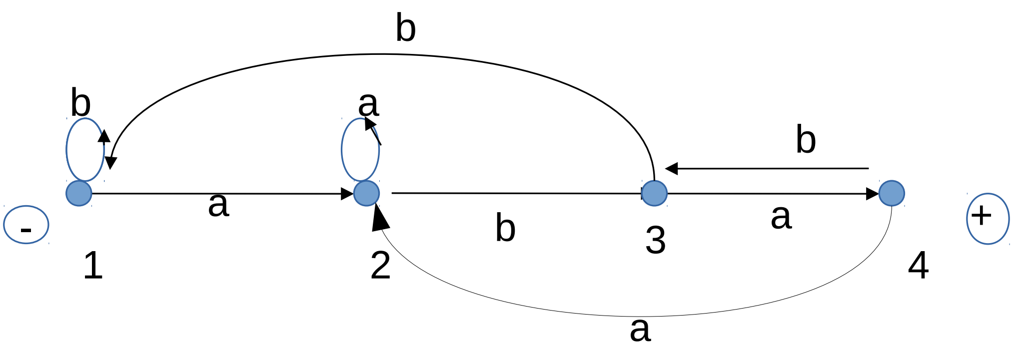
Exemple :

1 x bbaabab = 3

1X babbababa = 4

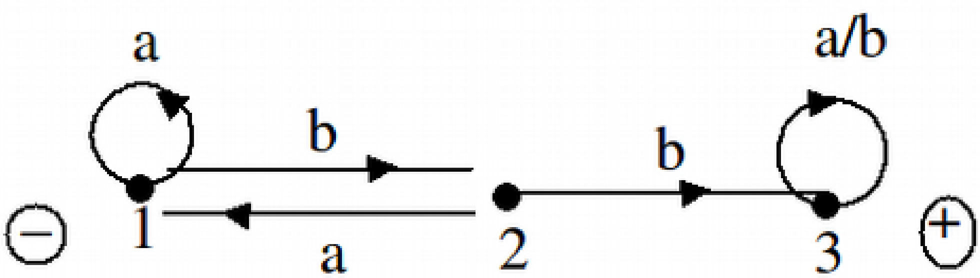
2 x aabbb = 1

A



6

Exercice :



1Xabba= ?
1Xaaabab= ?
2Xaab= ?

1Xu•bb•v= ? (u et v mots quelconques)

7

- **<u>Propriété 1:</u>**

Pour tout u, v ∈X*, σ*****(e,u • v)= σ*****(σ*****(e,u),v)

e x u • v =(e x u) x v

- *<u>Démonstration</u>*: on démontre par récurrence sur k la proposition suivante:
  - **P(k):** $\forall e \in E, \forall u,v \in X^*, |u| = k \Rightarrow e \times (u \cdot v) = (e \times u) \times v$

9

2 b) Langages reconnus par un automate déterministe:

- **<u>Définition3</u>**:

Soit $A=(X, E, e_0, F, \sigma)$ un automate. Le langage reconnu par **A** est le langage:

$$L(A)=\{ m \in X^* / \ e_0 \ x \ m \in F\}$$

10

Le chemin étiqueté:

Soit A =(X,E, e_0, F,σ) un automate, e un état de E et m=$a_1a_2\ldots a_k$ un mot de X*. On note:

$\Lambda(e,m)=(t_0 = e, a_1, t_1, a_2, \ldots a_i, t_i, \ldots a_k, t_k)$

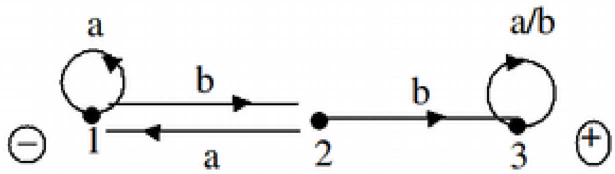
avec

$$\forall\, i \in \{0,1,\ldots,k\}, t_i \in E,$$

$$\forall\, i \in \{0,1,\ldots,k-1\}, t_i \times a_{i+1} = t_{i+1}$$

11

Exemple de chemin étiqueté :



$\Lambda(1,aabbabb)=(1,a,1,a,1,\underline{b,2,b},3,a,3,b,3,b,3)$

Première fois
dans l'état 3

$$1 \xrightarrow{a} 1 \xrightarrow{a} 1 \xrightarrow{b} 2 \xrightarrow{b} 3 \xrightarrow{a} 3 \xrightarrow{b} 3 \xrightarrow{b} 3$$

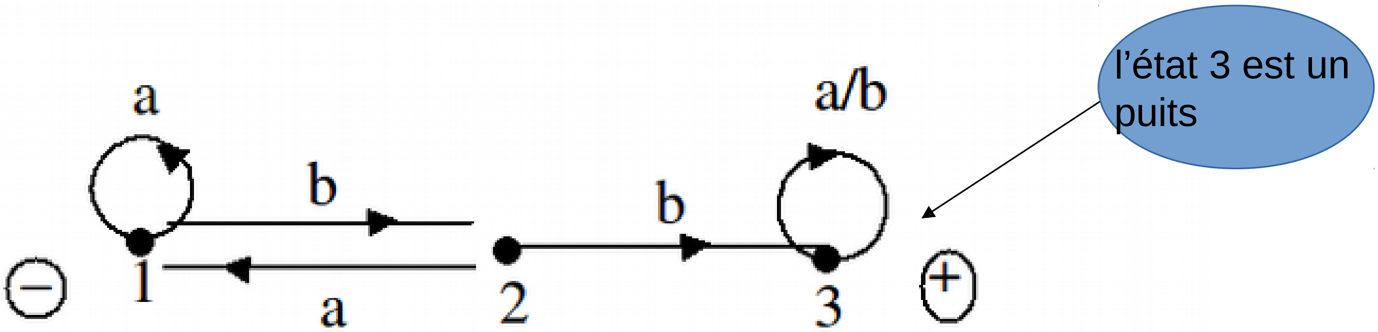
Démonstration de langage :

L(A) = {m∈X*,bb facteur de m} = {u•bb•v, u,v∈ X*} **?**

1) 1Xu•bb•v=3 ?

2) Si 1Xm=3 alors bb facteur de m ?

13



1Xu•bb•v= 3 ? (u et v mots quelconques)

Démonstration :

1er cas : 1xu=1 alors :

1xu•bb•v=(1xu)Xbb•v=(1xbb)xv=3xv=3

(dernière égalité : l'état 3 est un puits)

2ème cas : 1xu=2 alors :

1xu•bb•v=(1xu)Xbb•v=(2xbb)xv=3xv=3

3ème cas : 1xu=3 alors :

1xu•bb•v=(1xu)xbb•v=(3xbb)xv=3xv=3

Si 1Xm=3 alors bb facteur de m ?

Par hypothèse Λ(1,m) est de la forme :

$\Lambda(1,m)=(t_0=1,\ldots\ldots,b,t_{i-1}=2,b,t_i=3,\ldots,3,\ldots,3,\ldots,t_k=3)$

Donc bb facteur de m

(explications :

Par hypothèses : $t_0=1$ et $t_k=3$.

Notons i le plus petit indice tel que $t_i=3$

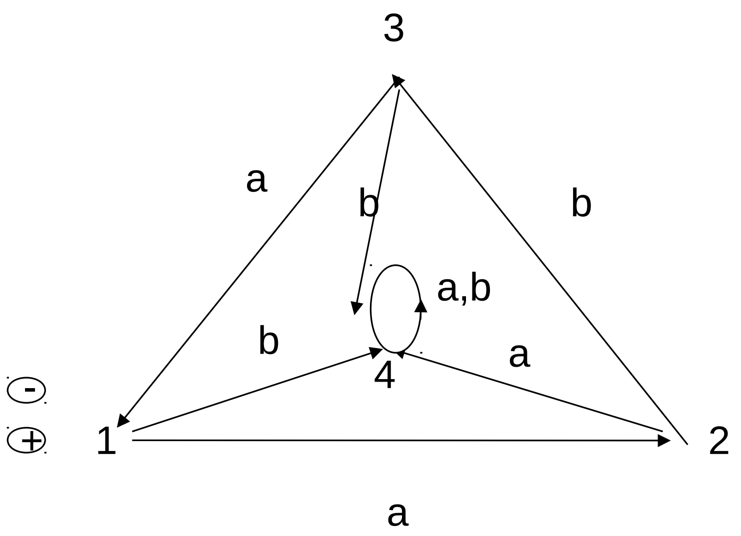
Par analyse « rétro » de l'automate :

$t_{i-1}=2$, $a_i=b$ et $a_{i-1}=b$

15

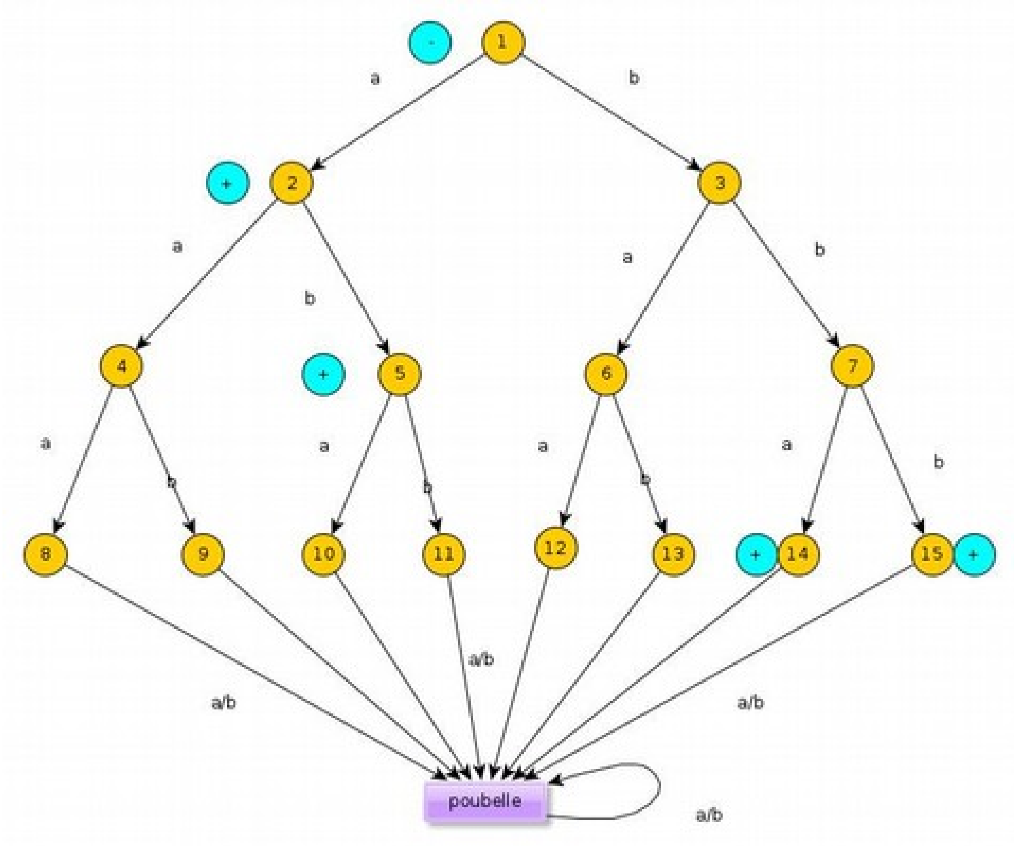
Exemple 2: A_2

- A_2



16

Automate reconnaissant un langage fini L(A)={a,ab,bba,bbb}



Automate reconnaissant le complémentaire de L(A)

Si

$A=(X,E,e_0,F,\sigma)$ est un automate reconnaissant L(A)

alors

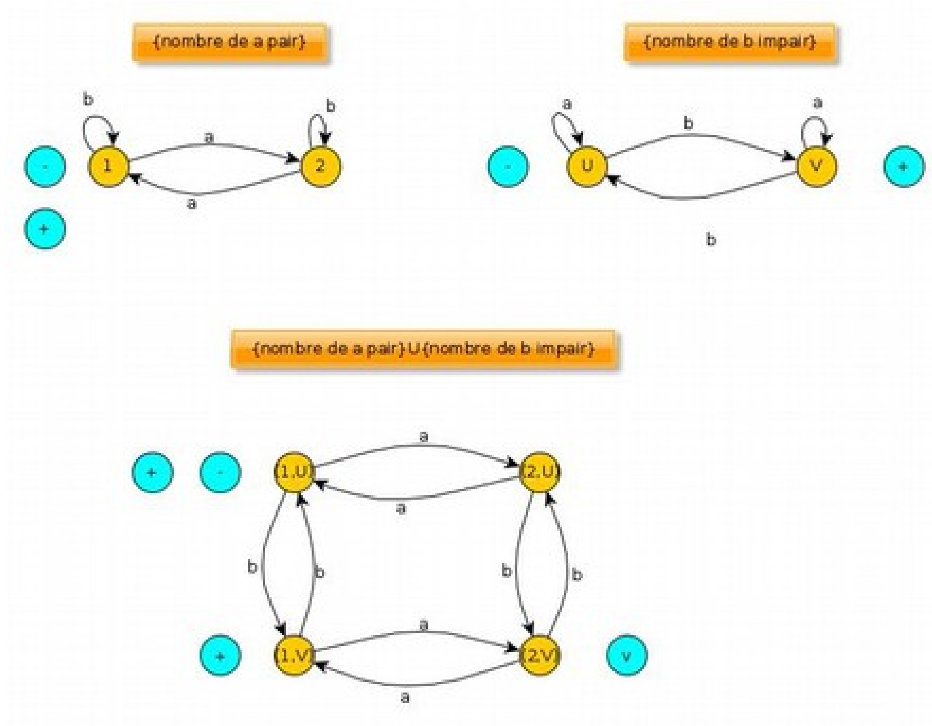
$A'=(X,E,e_0,E-F,\sigma)$ est un automate

reconnaissant le complémentaire de L(A)

(les états finaux de A' sont ceux qui ne sont pas finaux dans A)

3

Automate reconnaissant L(A) U L(A')



4

Automate reconnaissant l'intersection de L(A) et L(A')

La construction est la même que pour l'union mais les états finaux sont ceux qui sont finaux à la fois pour A et pour A' (ce sont les couples qui appartiennent à FXF')

5

Exercice 1 :

Dressez le diagramme de transition de l'automate reconnaissant tous les mots **n'ayant pas abb facteur** et uniquement ces mots.

6

Exercice 2

Dressez le diagramme de transition de l'automate reconnaissant tous les mots ayant **ab suffixe et le mot ε** (et uniquement ces mots).

9

Exemple: un automate reconnaissant les constantes réelles.

Les constantes réelles (en pascal par ex.):

2 -3.5 +2.5E-53 20E2 …

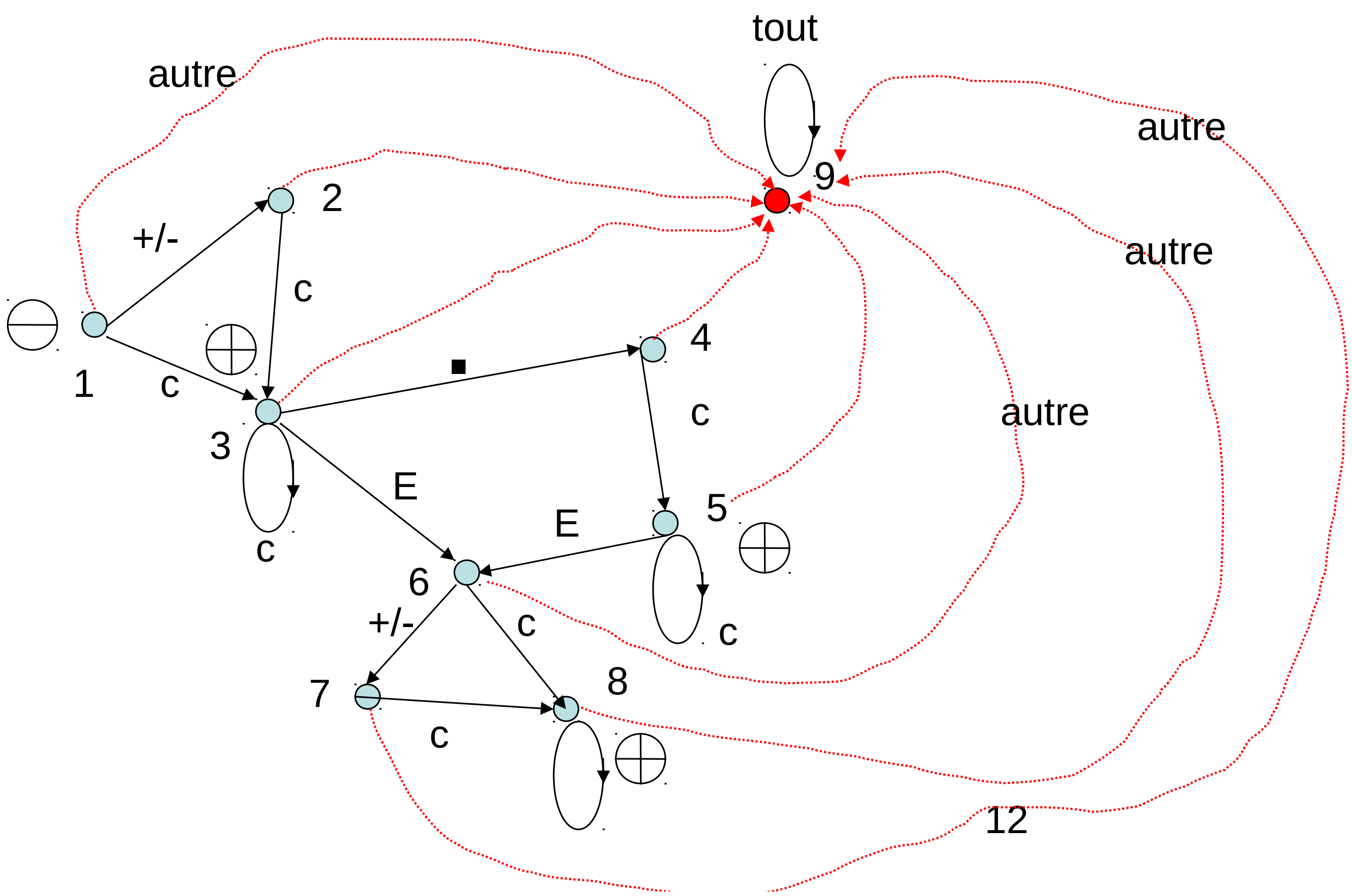
(pour simplifier on ne donnera pas de statut particulier à 0. On acceptera donc l'écriture 003.00E-005)

<u>L'alphabet est donc</u>:

X={0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ., E, +,-}

Les chiffres ayant le même statut, on notera plus simplement C

11



tout
autre
autre
autre
autre
9
2
+/-
c
1
c
3
c
■
4
c
E
E
5
c
6
+/-
c
7
8
c
12

2 C)Programmation des automates finis:

Algorithme ReconnaissanceParUnautomateDéterministe

Données: un automate $A=(X, E, e_0, F, \sigma)$ et un mot $m \in X^*$

Résultat: retourne vrai si $m \in L(A)$.

$e \leftarrow e_0$

$i \leftarrow 0$

tantque $i<|m|$ **faire**

$i \leftarrow i+1$

$e \leftarrow e \times m[i]$

retourner $(e \in F)$

13

2 d)Un langage non reconnaissable:

- **Langage des parenthèses** défini sur X={a,b}:

C'est un langage de X* noté **LP** défini par:

m∈LP ⇔ 1) $|m|_a = |m|_b$

2) Pour tout u *préfixe* de m on a

$$|u|_a \geq |u|_b$$

2 d)Un langage non reconnaissable:

- Propriété 2:

Il n'existe aucun automate A tel que LP = L(A).

15

Quelques mots de LP

Les mots ε, ab, aabb, aabbab, $a^n b^n$, aababaabbb

sont dans LP.

Le mot aaababbbabbaab n'est pas dans LP

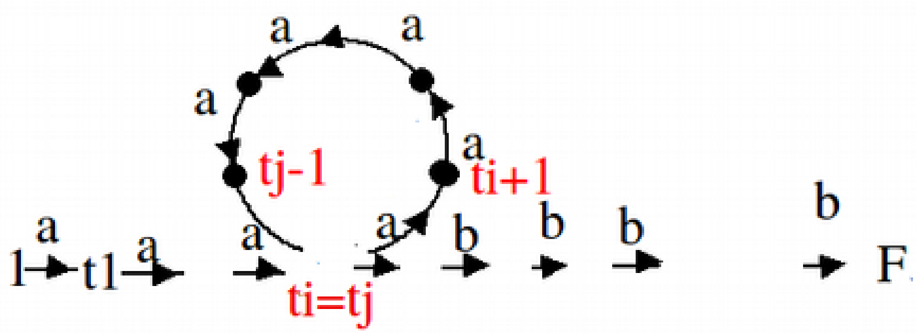
16

Idée de la démonstration :

Par l’absurde :

Si l’automate existait alors, pour tout entier n, 1 x $a^n b^n \in$ F.

Si n est supérieur au nombre d’état de l’automate:



Définition 4:

Les langages rationnels sont les langages reconnaissables par des automates finis.

(LP n’est donc pas rationnel)

18

Langages rationnels

Si L et L’ sont deux langages rationnels alors :

1) l’union de L et L’ est rationnel.
2) l’intersection de L et L’ est rationnel.
3) Le complémentaire de L est rationnel.
4) L•L’ est rationnel

19

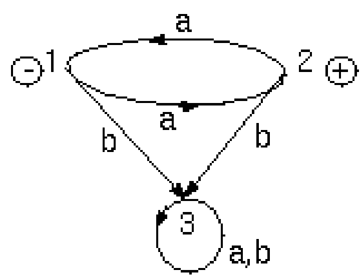
Exercices - Automates déterministes

Dans les exercices suivants, $X = \{a, b\}$.

Exercice 1
Soit $A = (X, E, e_0, F, \sigma)$ l'automate défini par : $X = \{a, b\}$, $E = \{1, 2, 3\}$, $e_0 = 1$, $F = \{3\}$, $\sigma(1, a) = 2$, $\sigma(1, b) = 1$, $\sigma(2, a) = 2$, $\sigma(2, b) = 3$, $\sigma(3, a) = 3$, $\sigma(3, b) = 3$.

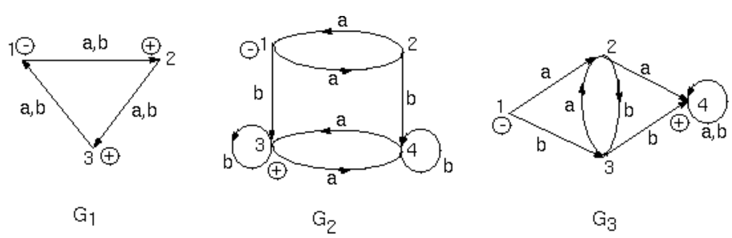
1. Donner la table et le graphe de transition de A.
2. Les mots *bbabb*, *aabaa*, *bbaaa* appartiennent-ils au langage $L(A)$?
3. Donner tous les mots de $L(A)$ de longueur inférieure ou égale à 3.
4. Déterminer $L(A)$. (Preuve)

Exercice 2
Soit A l'automate sur X dont le graphe de transition est



1. Déterminer X, E, e_0, F, σ tels que $A = (X, E, e_0, F, \sigma)$.
2. Donner tous les mots de $L(A)$ de longueur inférieure ou égale à 5.
3. Déterminer $L(A)$. (Preuve)

Exercice 3
Soit A_1, A_2, A_3 les automates de graphes de transition G_1, G_2, G_3.



Pour i de 1 à 3,

1. Donner les 7 premiers mots de $L(A_i)$ suivant l'ordre sur X^* : par longueur croissante et à longueur égale suivant l'ordre lexicographique.

2. Donner, si c'est possible, les 7 premiers mots de $L(A_i)$ suivant l'ordre lexicographique sur X^*.
3. Déterminer $L(A_i)$.

Exercice 4
Soit $X = \{a, b\}$. Pour i de 1 à 8, dessiner le graphe de transition G_i d'un automate A_i reconnaissant le langage L_i.

1. L_1 : ensemble des mots de X^* ayant aa comme sous-mot,
2. L_2 : ensemble des mots de X^* ayant aa comme facteur,
3. L_3 : ensemble des mots de X^* ayant aa comme préfixe,
4. L_4 : ensemble des mots de X^* ayant aa comme suffixe,
5. $L_5 = \{ab\}$,
6. $L_6 = \{(ab)^n, n \geq 0\}$,
7. $L_7 = \{m \in X^* \mid |m|_b \ mod \ 3 = 1\}$, ($L_7$ est l'ensemble des mots de X^* dont le nombre de b modulo 3 est égal à 1),
8. $L_8 = \{m \in X^* \mid |m|_b \ mod \ 3 \neq 1\}$,
9. $L_9 = \{m \in X^* \mid |m|_a$ est pair $\}$,
10. $L_{10} = L_7 \cap L_9$,
11. $L_{11} = L_7 \cup L_9$.

Exercice 5
Les langages suivants sont-ils rationnels ? (si oui, dessiner le graphe de transition d'un automate reconnaissant ce langage)
$L_1 = \{a^n b^p, n \geq 1, p \geq 1\}$,
$L_2 = \{a^n b^p, n \geq 0, p \geq 0\}$,
$L_3 = \{a^n b^n, n \geq 0\}$,
$L_4 = \{a^n b^p, n > p \geq 0\}$,
$L_5 = \{(ab)^n, n \geq 1\}$.

Exercice 6
On considère $A = (X, E, e_0, F, \sigma)$ l'automate défini par : $X = \{a, b\}$, $E = \{1, 2, 3, 4\}$, $e_0 = 1$, $F = \{4\}$, $\sigma(1, a) = 2$, $\sigma(1, b) = 1$, $\sigma(2, a) = 3$, $\sigma(2, b) = 1$, $\sigma(3, a) = 4$, $\sigma(3, b) = 1$, $\sigma(4, a) = 4$, $\sigma(4, b) = 4$.

1. (a) Donner sans justification, le graphe de transition de l'automate A.
   (b) Les mots a^2b^2a et ba^3b^2 appartiennent-ils au langage $L(A)$?
   (c) Donner sans justification les six premiers mots de $L(A)$ rangés
      i. par ordre de longueur croissante et à longueur égale suivant l'ordre lexicographique ;
      ii. par ordre lexicographique, si c'est possible.

(d) Donner sans justification, le graphe de transition d'un automate reconnaissant $L(A) \cup \{\varepsilon\}$.

2. Soit $L = \{m \in X^*, \exists u, v \in X^* \quad m = u \cdot a^3 \cdot v\}$.
   Montrer que $L(A) = L$.
3. Pour tout langage U sur X, on note U^* le langage sur X contenant les mots pouvant se factoriser en un nombre éventuellement nul de facteurs appartenant à U, c'est-à-dire

   $U^* = \{\varepsilon\} \cup \{m \in X^* \mid \exists k \in \mathbf{N}, k \geq 1, \exists m_1, m_2, ..., m_k \in U, m = m_1 \cdot m_2 \cdot ... \cdot m_k\}$

   Soit $L_1 = \{a^3, b\}$.

   (a) Donner sans justification les six premiers mots de L_1^* rangés
      i. par ordre de longueur croissante et à longueur égale suivant l'ordre lexicographique ;
      ii. par ordre lexicographique, si c'est possible.

   (b) Donner sans justification, le graphe de transition d'un automate reconnaissant L_1.

   (c) Donner sans justification, le graphe de transition d'un automate reconnaissant L_1^*.

   (d) Donner sans justification, le graphe de transition d'un automate reconnaissant $L_1^* \cap L$.

Exercice 7

On considère $A = (X, E, e_0, F, \sigma)$ l'automate défini par : $X = \{a, b\}$, $E = \{1, 2, 3\}$, $e_0 = 1$, $F = \{3\}$, $\sigma(1, a) = 2$, $\sigma(1, b) = 3$, $\sigma(2, a) = 2$, $\sigma(2, b) = 3$, $\sigma(3, a) = 3$, $\sigma(3, b) = 3$.

1. Donner sans justification, le graphe de transition de l'automate A.
2. Les mots a^2b^2a et a^4 appartiennent-ils au langage $L(A)$?
3. Donner sans justification, si c'est possible, les cinq premiers mots de $L(A)$ rangés
   (a) par ordre de longueur croissante et à longueur égale suivant l'ordre lexicographique ;
   (b) par ordre lexicographique.
4. Démontrer que le langage L_1 reconnu par l'automate A, est l'ensemble de tous les mots de X^* contenant au moins un b.
5. Donner sans justification, le graphe de transition d'un automate A_1 ayant moins d'états que A tel que $L(A_1) = L_1$.
6. Donner sans justification, le graphe de transition d'un automate reconnaissant le langage
   (a) $L_2 = L_1 \cup \{\varepsilon\}$;
   (b) $L_3 = L_1 \cup \{a\}$;
   (c) $L_4 = \{m \in X^*, |m| \text{ est pair }\}$;
   (d) $L_5 = L_1 \cap L_4$.