Automates, Langages et Logique

1. Alphabets – Mots – Langages

***L2**, Université d'Orléans — S1 2024/2025*

Nicolas Ollinger

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☀, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\frac{\sqrt{5}+1}{2}, \pi\right\}$

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☀, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\dfrac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, ...

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☀, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\frac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, … et les **lettres minuscules** de fin d'alphabet pour représenter les **lettres** : x, y, z, …

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☼, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\frac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, … et les **lettres minuscules** de fin d'alphabet pour représenter les **lettres** : x, y, z, …

Pour fabriquer des **mots**, on colle les lettres les unes derrière les autres !

★★★★★★

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☼, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\dfrac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, … et les **lettres minuscules** de fin d'alphabet pour représenter les **lettres** : x, y, z, …

Pour fabriquer des **mots**, on ~~colle~~**concatène** les lettres.

★★★★★★

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☀, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\frac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, ... et les **lettres minuscules** de fin d'alphabet pour représenter les **lettres** : x, y, z, ...

Pour fabriquer des **mots**, on ~~colle~~**concatène** les lettres.

★★★★★★

L'ordre d'apparition des lettres dans un mot est important.

Alphabet

Définition Un **alphabet** est un ensemble fini de **lettres**, ou **symboles**.

$\{0, 1\}$ $\{a, b, c\}$ {☀, ☁, 🌧}

{I, V, X, D, C, L, M} {★, ★, ★, ★} $\left\{\frac{\sqrt{5}+1}{2}, \pi\right\}$

Dans ce cours, on utilise les **lettres capitales** du début de l'alphabet pour représenter pour les **alphabets** : A, B, C, … et les **lettres minuscules** de fin d'alphabet pour représenter les **lettres** : x, y, z, …

Pour fabriquer des **mots**, on ~~colle~~**concatène** les lettres.

★★★★★★

L'ordre d'apparition des lettres dans un mot est ~~important~~**significatif**.

Mot

Définition Un **mot** u sur l'**alphabet** A est une suite finie $u_0 u_1 \cdots u_{n-1}$ de **lettres** de A. L'entier n est la **longueur** de u, notée $|u|$.

Mot

Définition Un **mot** u sur l'**alphabet** A est une suite finie $u_0 u_1 \cdots u_{n-1}$ de **lettres** de A. L'entier n est la **longueur** de u, notée $|u|$.

Le **plus petit** mot est le **mot vide**, de longueur 0, noté ε.

Mot

Définition Un **mot** u sur l'**alphabet** A est une suite finie $u_0 u_1 \cdots u_{n-1}$ de **lettres** de A. L'entier n est la **longueur** de u, notée $|u|$.

Le **plus petit** mot est le **mot vide**, de longueur 0, noté ε.

Dans ce cours, on **indexe les lettres** des mots en partant de 0. Ainsi u_i, ou encore $u[i]$, désigne la $i+1$ e lettre du mot u.

Mot

Définition Un **mot** u sur l'**alphabet** A est une suite finie $u_0 u_1 \cdots u_{n-1}$ de **lettres** de A. L'entier n est la **longueur** de u, notée $|u|$.

Le **plus petit** mot est le **mot vide**, de longueur 0, noté ε.

Dans ce cours, on **indexe les lettres** des mots en partant de 0. Ainsi u_i, ou encore $u[i]$, désigne la $i+1^{\text{e}}$ lettre du mot u.

Remarque Attention, **certains** indexent parfois à partir de 1.

Mot

Définition Un **mot** u sur l'**alphabet** A est une suite finie $u_0 u_1 \cdots u_{n-1}$ de **lettres** de A. L'entier n est la **longueur** de u, notée $|u|$.

Le **plus petit** mot est le **mot vide**, de longueur 0, noté ε.

Dans ce cours, on **indexe les lettres** des mots en partant de 0. Ainsi u_i, ou encore $u[i]$, désigne la $i+1$ᵉ lettre du mot u.

Remarque Attention, **certains** indexent parfois à partir de 1.

L'**ensemble de tous les mots** sur l'alphabet A est noté A^*.

$$\{a, b\}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, \ldots\}$$

Concaténation

Définition La **concaténation** de deux mots u et v sur l'alphabet A est le mot $u \cdot v$, parfois abrégé uv, de longueur $|u| + |v|$, obtenu en juxtaposant les deux mots.

$$\text{si } u = u_0 \cdots u_m$$
$$\text{et } v = v_0 \cdots v_n$$
$$\text{alors } u \cdot v = u_0 \cdots u_m v_0 \cdots v_n$$

Concaténation

Définition La **concaténation** de deux mots u et v sur l'alphabet A est le mot $u \cdot v$, parfois abrégé uv, de longueur $|u| + |v|$, obtenu en juxtaposant les deux mots.

$$\begin{aligned} \text{si } u &= u_0 \cdots u_m \\ \text{et } v &= v_0 \cdots v_n \\ \text{alors } u \cdot v &= u_0 \cdots u_m v_0 \cdots v_n \end{aligned}$$

On note u^n la puissance n^{e} de u, la **concaténation** de n copies de u.

$$\begin{aligned} u^0 &= \varepsilon \\ u^1 &= u \\ u^2 &= uu \\ &\vdots \\ u^n &= uu \cdots u \end{aligned}$$

Exercice (Fine et Wilf) Montrer que pour toute paire de mots $u, v \in A^*$, les deux affirmations suivantes sont équivalentes :

(1) les mots u et v commutent, *i.e.* $uv = vu$;

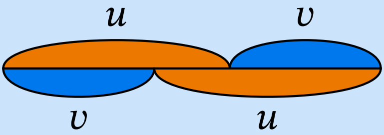
(2) il existe un mot $w \in A^*$ dont u et v sont des puissances, *i.e.* $\exists m, n \in \mathbb{N} \quad u = w^m \wedge v = w^n$.

Exercice (Fine et Wilf) Montrer que pour toute paire de mots $u, v \in A^*$, les deux affirmations suivantes sont équivalentes :

(1) les mots u et v commutent, *i.e.* $uv = vu$;

(2) il existe un mot $w \in A^*$ dont u et v sont des puissances, *i.e.* $\exists m, n \in \mathbb{N} \quad u = w^m \wedge v = w^n$.

Indice

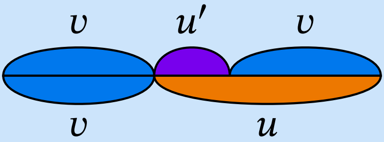


Exercice (Fine et Wilf) Montrer que pour toute paire de mots $u, v \in A^*$, les deux affirmations suivantes sont équivalentes :

(1) les mots u et v commutent, *i.e.* $uv = vu$;

(2) il existe un mot $w \in A^*$ dont u et v sont des puissances, *i.e.* $\exists m, n \in \mathbb{N} \quad u = w^m \wedge v = w^n$.

Indice

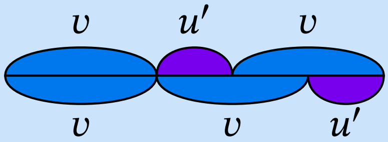


Exercice (Fine et Wilf) Montrer que pour toute paire de mots $u, v \in A^*$, les deux affirmations suivantes sont équivalentes :

(1) les mots u et v commutent, *i.e.* $uv = vu$;

(2) il existe un mot $w \in A^*$ dont u et v sont des puissances, *i.e.* $\exists m, n \in \mathbb{N} \quad u = w^m \wedge v = w^n$.

Indice



Facteurs

La **concaténation** compose les mots pour en former de nouveaux. On peut aussi **décomposer les mots**, les factoriser, en **facteurs**.

Facteurs

La **concaténation** compose les mots pour en former de nouveaux. On peut aussi **décomposer les mots**, les factoriser, en **facteurs**.

Définition Soient u, v et w trois mots de A^* :

1. le mot v est un **facteur** du mot uvw ;
2. le mot v est un **préfixe** du mot vw ;
3. le mot v est un **suffixe** du mot uv.

On ajoute le qualificatif **propre** lorsqu'on demande que les deux mots soient distincts.

Facteurs

La **concaténation** compose les mots pour en former de nouveaux. On peut aussi **décomposer les mots**, les factoriser, en **facteurs**.

Définition Soient u, v et w trois mots de A^* :

1. le mot v est un **facteur** du mot uvw ;
2. le mot v est un **préfixe** du mot vw ;
3. le mot v est un **suffixe** du mot uv.

On ajoute le qualificatif **propre** lorsqu'on demande que les deux mots soient distincts.

Le mot vide ε est **préfixe**, **suffixe** et **facteur** de chaque mot de A^*.

serttel te stoM

Définition L'**image miroir** u^R d'un mot u de A^* est le mot obtenu en lisant u de **droite à gauche**, *i.e.* si $u = u_0 \cdots u_{n-1}$ alors $u^R = u_{n-1} \cdots u_0$.

serttel te stoM

Définition L'**image miroir** u^R d'un mot u de A^* est le mot obtenu en lisant u de **droite à gauche**, *i.e.* si $u = u_0 \cdots u_{n-1}$ alors $u^R = u_{n-1} \cdots u_0$.

Un **palindrome** est un mot $u \in A^*$ égal à son image miroir : $u = u^R$.

Exemples `kayak` `mon␣nom` `roma␣summus␣amor` 선생생선

serttel te stoM

Définition L'**image miroir** u^R d'un mot u de A^* est le mot obtenu en lisant u de **droite à gauche**, *i.e.* si $u = u_0 \cdots u_{n-1}$ alors $u^R = u_{n-1} \cdots u_0$.

Un **palindrome** est un mot $u \in A^*$ égal à son image miroir : $u = u^R$.

Exemples `kayak` `mon␣nom` `roma␣summus␣amor` 선생생선

On note $|u|_x$ le **nombre d'occurrences** de $x \in A$ dans le mot $u \in A^*$.

Exercice (Carton) Soit (u_i) une suite de mots sur l'alphabet $A = \{a, b\}$ définie par récurrence par

$$\begin{cases} u_0 = a \\ u_1 = b \\ u_{n+2} = u_{n+1}u_n & \forall n \geqslant 0. \end{cases}$$

1. Montrer que ba est un **suffixe** de u_{2n} quand $n \geqslant 1$
   et que ab est un **suffixe** de u_{2n+1} quand $n \geqslant 1$;
2. Montrer que le mot v_n obtenu en supprimant les deux dernières lettres de u_{n+2} est toujours un **palindrome**.

Langage

Définition Un **langage** sur l'alphabet A est un **ensemble de mots** sur cet alphabet, $L \subseteq A^*$.

Exemple Les langages **vide** $\emptyset$ et **plein** A^*, le langage $\{a^n b^n | n \in \mathbb{N}\}$.

Langage

Définition Un **langage** sur l'alphabet A est un **ensemble de mots** sur cet alphabet, $L \subseteq A^*$.

Exemple Les langages **vide** $\emptyset$ et **plein** A^*, le langage $\{a^n b^n | n \in \mathbb{N}\}$.

Remarque Tout **langage** sur A est une **partie** de A^*.

Langage

Définition Un **langage** sur l'alphabet A est un **ensemble de mots** sur cet alphabet, $L \subseteq A^*$.

Exemple Les langages **vide** $\emptyset$ et **plein** A^*, le langage $\{a^n b^n | n \in \mathbb{N}\}$.

Remarque Tout **langage** sur A est une **partie** de A^*.

On étend la notion de **concaténation** aux langages.

Définition Soient L_1 et L_2 deux **langages** sur A. La **concaténation**, ou **produit**, de L_1 et L_2 est le langage $L_1 \cdot L_2$, aussi noté $L_1 L_2$, tel que

$$L_1 L_2 = \{uv | u \in L_1 \wedge v \in L_2\} \quad .$$

Opérations booléennes

On peut **combiner** les langages pour en créer de nouveaux.

Définition Soient L_1 et L_2 deux **langages** sur A. On définit les opérations booléennes suivantes :

$$L_1 + L_2 = L_1 \cup L_2 = \{u \in A^* | u \in L_1 \vee u \in L_2\} \qquad \textbf{(Union)}$$

$$L_1 \cap L_2 = \{u \in A^* | u \in L_1 \wedge u \in L_2\} \qquad \textbf{(Intersection)}$$

$$\overline{L_1} = A^* \setminus L_1 = \{u \in A^* | u \notin L_1\} \qquad \textbf{(Complément)}$$

$$L_1 \setminus L_2 = \{u \in A^* | u \in L_1 \wedge u \notin L_2\} \qquad \textbf{(Différence)}$$

$\emptyset$ est l'**élément neutre** pour l'**union** : $L + \emptyset = \emptyset + L = L \qquad \forall L \subseteq A^*$.

$\{\varepsilon\}$ est l'**élément neutre** pour le **produit** : $L\{\varepsilon\} = \{\varepsilon\}L = L \qquad \forall L \subseteq A^*$.

Exercice Soient A un alphabet, L le langage des mots de longueur paire de A^* et L' le langage des mots de longueur impaire de A^*. Calculer les langages suivants : $L + L'$, LL', LL, $L'L$, $L'L'$, $\overline{L}$, $\overline{L'}$.

Étoile

On note L^n la puissance n^{e} de L, la **concaténation** de n copies de L.

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^{n+1} = LL^n & \forall n \in \mathbb{N} \end{cases}$$

Remarque Attention, en général L^n est différent de $\{u^n | u \in L\}$!

Étoile

On note L^n la puissance n^e de L, la **concaténation** de n copies de L.

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^{n+1} = LL^n \qquad \forall n \in \mathbb{N} \end{cases}$$

Remarque Attention, en général L^n est différent de $\{u^n | u \in L\}$!

Définition L'**étoile** d'un langage L sur A est le langage $L^* = \cup_{n \geqslant 0} L^n$.

La notation A^* pour le **langage plein** est compatible avec cette notation en considérent A comme le langage des **mots à une lettre**.

Étoile

On note L^n la puissance n^{e} de L, la **concaténation** de n copies de L.

$$\begin{cases} L^0 = \{\varepsilon\} \\ L^{n+1} = LL^n \qquad \forall n \in \mathbb{N} \end{cases}$$

Remarque Attention, en général L^n est différent de $\{u^n | u \in L\}$!

Définition L'**étoile** d'un langage L sur A est le langage $L^* = \cup_{n \geqslant 0} L^n$.

La notation A^* pour le **langage plein** est compatible avec cette notation en considérent A comme le langage des **mots à une lettre**.

On note L^+ le langage $LL^* = \cup_{n>0} L^n$.

Exercice Décrire l'étoile de chacun des langages suivants :

$$
\begin{aligned}
L_0 &= \{a\} \\
L_1 &= \{a, ba\} \\
L_2 &= \{a^n b | n \in \mathbb{N}\} \\
L_3 &= \emptyset
\end{aligned}
$$

Quotient

L'opération de **quotient** sera très utile pour étudier les automates.

Définition Le **quotient gauche** d'un langage $L \subseteq A^*$ par un mot $u \in A^*$ est le langage $u^{-1}L = \{v \in A^* | uv \in L\}$.

Définition Le **quotient droit** d'un langage $L \subseteq A^*$ par un mot $u \in A^*$ est le langage $Lu^{-1} = \{v \in A^* | vu \in L\}$.

Ces notations sont **étendues aux langages** par

$$K^{-1}L = \{v \in A^* | \exists u \in K \quad uv \in L\}$$
$$LK^{-1} = \{v \in A^* | \exists u \in K \quad vu \in L\}$$

Monoïde

Définition Un **monoïde** est un ensemble E muni d'une **operation interne** $\cdot$ **associative** et d'un élément neutre $1 \in E$, *i.e.*, pour tout $x, y, z \in E$,

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

$$x \cdot 1 = 1 \cdot x = x$$

Le langage A^* muni de la concaténation et du mot vide ε est un **monoïde**.

Un **sous-monoïde** de A^* est un langage **stable par concaténation**.

Le **plus petit** sous-monoïde contenant $L \subseteq A^*$ est L^*.

Décrire un langage

Énumération Lister les éléments du langage

$$L = \{\varepsilon, a, aa, ab, ba, aaa, aab, aba, baa, bab, \cdots\}$$

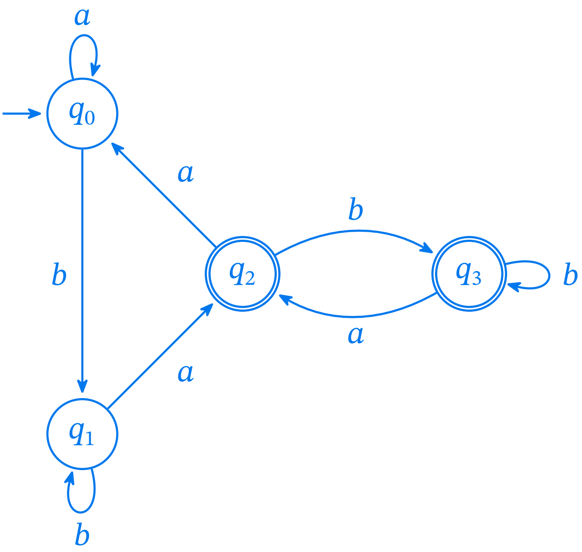
En langue naturelle Décrire des mots avec des mots

« Les mots sur l'alphabet $\{a, b\}$ qui contiennent deux fois plus de a que de b. »

Définition/formule À l'aide d'une formule mathématique

$$L = \{u \in \{0,1\}^* \mid |u|_0 < 3|u|_1 + 2\}$$

Décrire un langage : automates finis



Décrire un langage : expressions rationnelles

Une **expression rationnelle** décrit un langage à partir du langage vide $\emptyset$, des langages à une lettre $\{x\}$ et des opérations de **produit**, **somme** et **étoile**.

La syntaxe est simplifiée pour une écriture plus lisible :

$$E, F \coloneqq \emptyset \mid x \mid EF \mid E + F \mid E^* \mid (E)$$

$$(a+b)^* ab (a+b)^* ba (a+b)^*$$

En pratique

```
$ grep '[bl].*fr.g$' < /usr/share/dict/words
bullfrog
leapfrog
```

grep affiche les lignes qui appartiennent au langage.

`[bl]` `.*` `f` `r` `.` `g`

« Les **chaînes** contenant `b` ou `l` puis, plus loin, `fr`, une lettre, puis `g`. »