IF04M050 : Théorie des langages

Alphabets et langages

Jérôme Delobelle
jerome.delobelle@u-paris.fr

LIPADE - Université de Paris

Alphabets et langages

1. Alphabets
2. Opérations sur les mots
3. Monoïde
4. Langages
5. Expressions régulières

Alphabets

Alphabet

Alphabet

Un **alphabet** est un ensemble fini de symboles.

Exemples :

- $A = \{0, 1\}$
- $\Sigma = \{a, b, c\}$
- $\Theta = \{if, then, else, a, b\}$
- $F = \{\rightarrow, \leftarrow, \uparrow, \downarrow\}$

Mot

Mot

Un **mot** sur l'alphabet X est une séquence finie et ordonnée, éventuellement vide, d'éléments de l'alphabet.
C'est une suite de symboles.

Par exemple, *abbac* et *ba* sont deux mots de l'alphabet $\{a, b, c\}$.

Mot

Mot

Un **mot** sur l'alphabet X est une séquence finie et ordonnée, éventuellement vide, d'éléments de l'alphabet.
C'est une suite de symboles.

Par exemple, *abbac* et *ba* sont deux mots de l'alphabet $\{a, b, c\}$.

Mot vide

Le **mot vide**, noté ϵ, correspond à la suite de symboles vide.

Longueur d'un mot

Longueur d'un mot

La **longueur** d'un mot w est le nombre de symboles constituant ce mot. On la note $|w|$.

Le mot vide est de longueur 0.

Par exemple, $|abbac| = 5$, $|ba| = 2$ et $|\epsilon| = 0$.

Longueur d'un mot

Longueur d'un mot

La **longueur** d'un mot w est le nombre de symboles constituant ce mot. On la note $|w|$.

Le mot vide est de longueur 0.

Par exemple, $|abbac| = 5$, $|ba| = 2$ et $|\epsilon| = 0$.

Ensemble de mots

L'**ensemble de mots** sur un alphabet X est noté X^* (fermeture transitive).

Par exemple, si $X = \{a, b, c\}$,
$X^* = \{\epsilon, a, b, c, aa, ab, ac, ba, bb, bc, ca, cb, cc, aaa, aab, ...\}$

Notations

Soit $w \in X^*$

- $|w|$ est la longueur de w
- X est l'alphabet
- x est un symbole de X
- $|w|_x$ est le **nombre d'occurences** de x dans w.

Exemple, avec $X = \{a, b\}$

- $|abb|_a = 1$
- $|abb|_b = 2$
- $|abb| = |abb|_a + |abb|_b = 3$

Opérations sur les mots

Concaténation (produit)

Concaténation (produit) de symboles

Soit un alphabet X et $w_1, w_2 \in X^*$ tels que

$$\begin{array}{rcll} w_1 & = & a_1a_2a_3\ldots a_n & \forall i \in \{1,\ldots,n\}, a_i \in X \\ w_2 & = & b_1b_2b_3\ldots b_p & \forall i \in \{1,\ldots,p\}, b_i \in X \end{array}$$

w_1 et w_2 sont des **concaténation de symboles** (mots).

Concaténation (produit) de mots

Soit un alphabet X et $w_1, w_2 \in X^*$ des concaténation de symboles.

Alors w tel que :

$$w = w_1.w_2 = a_1a_2a_3\ldots a_nb_1b_2b_3\ldots b_p$$

est une **concaténation de mots**.

Concaténation (produit)

Propriétés

- Le produit est associatif

$$\forall w_1, w_2, w_3 \in X^*, \quad \begin{aligned} w_1.(w_2.w_3) &= (w_1.w_2).w_3 \\ &= w_1.w_2.w_3 \end{aligned}$$

- ϵ est l'élément neutre du produit

$$\forall w \in X^*,\ \epsilon.w = w.\epsilon = w$$

- $\forall w, z \in X^*,\ |w.z| = |w| + |z|$
- Le produit **n'est pas** commutatif

Puissance

Puissance

Soit un alphabet X et $w \in X^*$.

$$w^n = \begin{cases} \epsilon & \text{si } n = 0 \\ w.w^{n-1} & \text{si } n > 0 \end{cases}$$

Puissance

Soit un alphabet X et $w \in X^*$.

$$w^n = \begin{cases} \epsilon & \text{si } n = 0 \\ w.w^{n-1} & \text{si } n > 0 \end{cases}$$

Par exemple, soit $X = \{a, b\}$ et $w = abb$

- $w^0 = \epsilon$
- $w^1 = abb$
- $w^2 = w.w = abbabb$
- $w^3 = w.w^2 = abbabbabb$

Egalité

Egalité de deux mots

Deux mots sont **égaux** s'ils sont de même longueur et s'ils ont des lettres identiques de positionnements identiques.

Soit un alphabet X et $w_1, w_2 \in X^*$ tels que

$$\begin{aligned} w_1 &= a_1 a_2 a_3 \ldots a_n \quad \forall i \in \{1, \ldots, n\}, a_i \in X \\ w_2 &= b_1 b_2 b_3 \ldots b_p \quad \forall i \in \{1, \ldots, p\}, b_i \in X \end{aligned}$$

On a $w_1 = w_2$ si et seulement si $p = n$ et $\forall i \in [1, n]$, $a_i = b_i$.

Préfixe et suffixe

Préfixe, suffixe

Soit un alphabet X et $w, u \in X^*$.

u est un **préfixe** de w si et seulement si $\exists v \in X^*$ tel que $w = u.v$

Préfixe et suffixe

Préfixe, suffixe

Soit un alphabet X et $w, u \in X^*$.

u est un **préfixe** de w si et seulement si $\exists v \in X^*$ tel que $w = u.v$

u est un **suffixe** de w si et seulement si $\exists v \in X^*$ tel que $w = v.u$

Préfixe et suffixe

Préfixe, suffixe

Soit un alphabet X et $w, u \in X^*$.

u est un **préfixe** de w si et seulement si $\exists v \in X^*$ tel que $w = u.v$

u est un **suffixe** de w si et seulement si $\exists v \in X^*$ tel que $w = v.u$

Soit $X = \{a, b\}$, $w = babb$

- Les préfixes de w sont ϵ, b, ba, bab, $babb$
- Les suffixes de w sont ϵ, b, bb, abb, $babb$

Préfixe et suffixe propres

Préfixe et suffixe propres

Soit un alphabet X et $w, u \in X^*$.

u est un **préfixe propre** de w si et seulement si u est un préfixe de w et u est différent de w.

u est un **suffixe propre** de w si et seulement si u est un suffixe de w et u est différent de w.

Préfixe et suffixe propres

Préfixe et suffixe propres

Soit un alphabet X et $w, u \in X^*$.

u est un **préfixe propre** de w si et seulement si u est un préfixe de w et u est différent de w.

u est un **suffixe propre** de w si et seulement si u est un suffixe de w et u est différent de w.

Soit $X = \{a, b\}$, $w = babb$

- Les préfixes propres de w sont ϵ, b, ba, bab
- Les suffixes propres de w sont ϵ, b, bb, abb

Miroir d'un mot

Miroir d'un mot

Soit un alphabet X et $w \in X^*$ tel que $w = a_1 a_2 a_3 \ldots a_n$, avec $\forall i \in \{1, \ldots, n\}$, $a_i \in X$.

Le **miroir** de w, noté $\tilde{w}$, est défini par

$$\tilde{w} = a_n a_{n-1} \ldots a_2 a_1$$

Miroir d'un mot

Miroir d'un mot

Soit un alphabet X et $w \in X^*$ tel que $w = a_1 a_2 a_3 \dots a_n$, avec $\forall i \in \{1, \dots, n\}$, $a_i \in X$.

Le **miroir** de w, noté $\tilde{w}$, est défini par

$$\tilde{w} = a_n a_{n-1} \dots a_2 a_1$$

Définition récursive :

$$\tilde{w} = \begin{cases} w & \text{si } w = \epsilon \\ \tilde{u}.a & \text{si } w = a.u, \text{avec } a \in X \end{cases}$$

Monoïde

Monoïde

Monoïde

$\langle E, \oplus, \epsilon \rangle$ est un monoïde si et seulement si :

- E est un ensemble
- $\oplus$ est une loi de composition interne sur E
- $\oplus$ est associative
- $\oplus$ possède un élément neutre $\epsilon \in E$

Monoïde

Monoïde

$\langle E, \oplus, \epsilon \rangle$ est un monoïde si et seulement si :

- E est un ensemble
- $\oplus$ est une loi de composition interne sur E
- $\oplus$ est associative
- $\oplus$ possède un élément neutre $\epsilon \in E$

Exemples de monoïdes :

- $\langle \mathbb{N}, +, 0 \rangle$
- $\langle \mathbb{N}, \times, 1 \rangle$
- $\langle X^*, ., \epsilon \rangle$: l'ensemble des mots sur l'alphabet X muni de l'opération de concaténation est un monoïde.

Sous-monoïde

Sous-monoïde

$\langle E', \oplus, \epsilon \rangle$ est un sous-monoïde de $\langle E, \oplus, \epsilon \rangle$ si et seulement si :

- E' est un sous-ensemble de E (i.e. $E' \subseteq E$)
- $\oplus$ est une loi de composition interne sur E'
- $\epsilon \in E'$

Sous-monoïde

Sous-monoïde

$\langle E', \oplus, \epsilon \rangle$ est un sous-monoïde de $\langle E, \oplus, \epsilon \rangle$ si et seulement si :

- E' est un sous-ensemble de E (i.e. $E' \subseteq E$)
- $\oplus$ est une loi de composition interne sur E'
- $\epsilon \in E'$

Pour montrer que M' est un sous-monoïde de M, il suffit de montrer que

1. l'élément neutre de M appartient à M'
2. la loi de composition interne est stable pour E' :

$$\forall x, y \in E', x.y \in E'$$

Monoïde libre

Ensemble de générateurs

Soit $M = \langle E, \oplus, \epsilon \rangle$ un monoïde. Un **ensemble de générateurs** de M est un sous ensemble E_1, avec $E_1 \subset E$, tel que tout élément de E, sauf l'élément neutre, est exprimable à l'aide d'une composition de E_1.

Monoïde libre

Ensemble de générateurs

Soit $M = \langle E, \oplus, \epsilon \rangle$ un monoïde. Un **ensemble de générateurs** de M est un sous ensemble E_1, avec $E_1 \subset E$, tel que tout élément de E, sauf l'élément neutre, est exprimable à l'aide d'une composition de E_1.

Exemple :

- $\{1\}$ est un générateur de $\langle \mathbb{N}, +, 0 \rangle$
  $\rightarrow$ Tout entier peut être exprimé comme une somme de 1
- L'ensemble des nombres premiers est un générateur de $\langle \mathbb{N}^*, \times, 1 \rangle$
  $\rightarrow$ Tout entier non nul peut être exprimé comme un produit de nombre premiers

Monoïde libre

Ensemble de générateurs indépendants

Soit $M = \langle E, \oplus, \epsilon \rangle$ un monoïde. Un **ensemble de générateurs indépendants** de M est un ensemble de générateurs tels que tout élément de E sauf l'élément neutre est exprimable *d'une et d'une seule façon* sous forme d'une composition de générateurs.

Monoïde libre

Ensemble de générateurs indépendants

Soit $M = \langle E, \oplus, \epsilon \rangle$ un monoïde. Un **ensemble de générateurs indépendants** de M est un ensemble de générateurs tels que tout élément de E sauf l'élément neutre est exprimable *d'une et d'une seule façon* sous forme d'une composition de générateurs.

Exemple :

- $\{1\}$ est un générateur indépendant de $\langle \mathbb{N}, +, 0 \rangle$
  - $\rightarrow$ Tout entier peut être exprimé d'une et d'une seule façon comme une somme de 1
- L'ensemble des nombres premiers n'est pas un générateur indépendant de $\langle \mathbb{N}^*, \times, 1 \rangle$
  - $\rightarrow$ Tout entier peut être exprimé comme un produit de nombre premiers, mais il y a plusieurs décompositions possibles. Par exemple, $12 = 2 * 3 * 2 = 2 * 2 * 3$.

Monoïde libre

Monoïde libre

Un monoïde possédant un ensemble de générateurs indépendants G sera dit **libre** et sera noté G^*.

Monoïde libre

Monoïde libre

Un monoïde possédant un ensemble de générateurs indépendants G sera dit **libre** et sera noté G^*.

Soit X un alphabet. Le monoïde $\langle X^*, ., \epsilon \rangle$ est un monoïde libre.

Langages

Langage

Langage

Un **langage** sur un alphabet X est une partie de X^*. C'est donc un ensemble de mots.

$$L \subset X^* \text{ où } L \in \mathcal{P}(X^*)$$

Langage

Langage

Un **langage** sur un alphabet X est une partie de X^*. C'est donc un ensemble de mots.

$$L \subset X^* \text{ où } L \in \mathcal{P}(X^*)$$

Soit $X = \{a, b\}$ un alphabet.

- $\emptyset$ est un langage
- $\{\epsilon\}$ est un langage
- $\{a, ba, bba\}$ est un langage
- $\{a^n \mid n \in \mathbb{N}\}$ est un langage — écriture en comprehension
- $\{\epsilon, a, aa, aaa, \ldots\}$ est un langage — écriture en extension

Opérations sur les langages

- Union : $A, B \subseteq X^*$, $A \cup B = \{w \in X^* | w \in A \text{ ou } w \in B\}$
  - Associative
  - Commutative
  - Elément neutre : ensemble vide $\emptyset$
  - Notée $+$ dans la théorie des langages
- Intersection : $A, B \subseteq X^*$, $A \cap B = \{w \in X^* | w \in A \text{ et } w \in B\}$
  - Associative
  - Commutative
  - Elément neutre X^*
- Différence : $A, B \subseteq X^*$, $A \setminus B = \{w \in X^* | w \in A \text{ et } w \notin B\}$
- Complémentaire : $A \subseteq X^*$, $\overline{A} = X^* \setminus A = \{w \in X^* | w \notin A\}$

Opérations sur les langages

Egalité de langages

Deux langages $A, B \subseteq X^*$ sont **égaux**, noté $A = B$, si et seulement si $A \subseteq B$ et $B \subseteq A$.

Opérations sur les langages

Egalité de langages

Deux langages $A, B \subseteq X^*$ sont **égaux**, noté $A = B$, si et seulement si $A \subseteq B$ et $B \subseteq A$.

Produit de langages

Soit deux langages $A, B \subseteq X^*$. Le **produit** de A et B est noté $A \circ B = \{u.v \mid u \in A \text{ et } v \in B\}$.

Opérations sur les langages

Egalité de langages

Deux langages $A, B \subseteq X^*$ sont **égaux**, noté $A = B$, si et seulement si $A \subseteq B$ et $B \subseteq A$.

Produit de langages

Soit deux langages $A, B \subseteq X^*$. Le **produit** de A et B est noté $A \circ B = \{u.v \mid u \in A \text{ et } v \in B\}$.

Attention !

- $\circ$ produit de langages
- $.$ produit de mots

$\Rightarrow$ Par la suite, nous les noterons de la même façon, le contexte fera la différence.

Opérations sur les langages

Théorème

Le produit de langages est distributif par rapport à l'union.

$$\forall A, B, C \subseteq X^* \quad \begin{array}{rcl} A.(B \cup C) & = & (A.B) \cup (A.C) \\ (B \cup C).A & = & (B.A) \cup (C.A) \end{array}$$

Opérations sur les langages

Théorème

Le produit de langages est distributif par rapport à l'union.

$$\begin{array}{rrcl} \forall A, B, C \subseteq X^* & A.(B \cup C) & = & (A.B) \cup (A.C) \\ & (B \cup C).A & = & (B.A) \cup (C.A) \end{array}$$

Ce théorème reste vrai pour des unions infinies

$$\begin{array}{rrcl} \forall A, B_i \subseteq X^* & A.(\bigcup_{i=1}^{\infty} B_i) & = & \bigcup_{i=1}^{\infty}(A.B_i) \\ & (\bigcup_{i=1}^{\infty} B_i).A & = & \bigcup_{i=1}^{\infty}(B_i.A) \end{array}$$

Opérations sur les langages

Théorème

Le produit de langages est distributif par rapport à l'union.

$$\forall A, B, C \subseteq X^* \quad \begin{array}{rcl} A.(B \cup C) & = & (A.B) \cup (A.C) \\ (B \cup C).A & = & (B.A) \cup (C.A) \end{array}$$

Ce théorème reste vrai pour des unions infinies

$$\forall A, B_i \subseteq X^* \quad \begin{array}{rcl} A.(\bigcup_{i=1}^{\infty} B_i) & = & \bigcup_{i=1}^{\infty}(A.B_i) \\ (\bigcup_{i=1}^{\infty} B_i).A & = & \bigcup_{i=1}^{\infty}(B_i.A) \end{array}$$

Attention ! Le produit de langages *n'est pas* distributif par rapport à l'intersection.

$$\forall A, B, C \subseteq X^*, A.(B \cap C) \subseteq (A.B) \cap (A.C)$$

Opérations sur les langages

Fermeture de Kleene

Soit $A \subseteq X^*$. On note $A^* = \bigcup_{i=0}^{\infty} A^i$ l'**opération étoile** (ou fermeture par étoile, ou fermeture de Kleene, ou fermeture itérative) du langage A.

Note : Comme $A^0 = \{\epsilon\}$, on a toujours $\epsilon \in A^*$

Opérations sur les langages

Fermeture de Kleene

Soit $A \subseteq X^*$. On note $A^* = \bigcup_{i=0}^{\infty} A^i$ l'**opération étoile** (ou fermeture par étoile, ou fermeture de Kleene, ou fermeture itérative) du langage A.

Note : Comme $A^0 = \{\epsilon\}$, on a toujours $\epsilon \in A^*$

Fermeture positive

Soit $A \subseteq X^*$. On note $A^+ = \bigcup_{i=1}^{\infty} A^i$ la **fermeture positive** du langage A.

Fermeture de Kleene

Soit $A \subseteq X^*$. On note $A^* = \bigcup_{i=0}^{\infty} A^i$ l'**opération étoile** (ou fermeture par étoile, ou fermeture de Kleene, ou fermeture itérative) du langage A.

Note : Comme $A^0 = \{\epsilon\}$, on a toujours $\epsilon \in A^*$

Fermeture positive

Soit $A \subseteq X^*$. On note $A^+ = \bigcup_{i=1}^{\infty} A^i$ la **fermeture positive** du langage A.

Théorème

Soit $A \subseteq X^*$. On a $A^+ = A.A^* = A^*.A$

Opérations sur les langages

Opération miroir

Soit $A \subseteq X^*$. On définit l'**opération miroir** comme étant :

$$A^R = \{\tilde{w} | w \in A\}$$

Opérations sur les langages

Opération miroir

Soit $A \subseteq X^*$. On définit l'**opération miroir** comme étant :

$$A^R = \{\tilde{w} | w \in A\}$$

Théorème

Soit $A, B \subseteq X^*$. On a $(A.B)^R = B^R.A^R$

Expressions régulières

Expressions régulières

Manière de décrire des motifs (*pattern* en anglais) représentant un ensemble de mots d'un langage donné.

Expressions régulières

Manière de décrire des motifs (*pattern* en anglais) représentant un ensemble de mots d'un langage donné.

Expressions régulières

Les **expressions régulières** sur un alphabet X sont formées à partir des règles suivantes :

1. $\emptyset$ et ϵ sont des expressions régulières
2. $\forall a \in X$, a est une expression régulière
3. Si α et β sont des expressions régulières alors

$$\left.\begin{array}{l}(\alpha + \beta) \\ (\alpha . \beta) \\ (\alpha)^*\end{array}\right\} \text{sont des expressions régulières}$$

Expressions régulières

Manière de décrire des motifs (*pattern* en anglais) représentant un ensemble de mots d'un langage donné.

Expressions régulières

Les **expressions régulières** sur un alphabet X sont formées à partir des règles suivantes :

1. $\emptyset$ et ϵ sont des expressions régulières
2. $\forall a \in X$, a est une expression régulière
3. Si α et β sont des expressions régulières alors

$$\left.\begin{array}{l}(\alpha+\beta)\\(\alpha.\beta)\\(\alpha)^*\end{array}\right\}\text{ sont des expressions régulières}$$

Priorité dans l'ordre décroissant : *, ., $+$

Expressions régulières

Manière de décrire des motifs (*pattern* en anglais) représentant un ensemble de mots d'un langage donné.

Expressions régulières

Les **expressions régulières** sur un alphabet X sont formées à partir des règles suivantes :

1. $\emptyset$ et ϵ sont des expressions régulières
2. $\forall a \in X$, a est une expression régulière
3. Si α et β sont des expressions régulières alors

$$\left.\begin{array}{l}(\alpha+\beta)\\(\alpha.\beta)\\(\alpha)^*\end{array}\right\}\text{ sont des expressions régulières}$$

Priorité dans l'ordre décroissant : *, ., $+$

$$a + bc^* \equiv (a + (b(c^*)))$$

Expressions régulières : Exemples

Soit l'alphabet $X = \{a, b, c\}$. Par exemple,

- ϵ
- c
- a^*
- $(a + b)^*$
- $ab(a + c)^* b^* (a + \epsilon)$

sont des expressions régulières sur X.

Exemples d'extensions des notations

Soit un alphabet X avec $a, b, c \in X$ et α une exp. régulière sur X.

Notation	Equivalence	Signification
$[abc]$	$a + b + c$	un symbole parmi un ens. de symboles
$[\hat{}abc]$	$X \backslash \{a, b, c\}$	tous les symboles de X sauf ...
$[a - z]$	$a + b + \cdots + y + z$	n'importe quelle lettre
$[0 - 9]$	$0 + \cdots + 9$	n'importe quel chiffre
$\alpha^?$	$\epsilon + \alpha$	0 ou une fois α
α^+	$\alpha\alpha^*$	au moins une fois α
$\alpha\{n\}$	α^n	α répétée exactement n fois
$\alpha\{n,\}$	$\alpha^n\alpha^*$	α répétée au moins n fois
$\alpha\{,m\}$	$\epsilon + \alpha + \alpha^2 + \cdots + \alpha^m$	α répétée au plus m fois
$\alpha\{n, m\}$	$\alpha^n + \alpha^{n+1} + \cdots + \alpha^m$	α répétée entre n et m fois ($n \leq m$)
$\hat{}\alpha$	αX^*	α doit apparaître en début de ligne
$\alpha\$$	$X^*\alpha$	α doit apparaître en fin de ligne

Exemples d'extensions des notations

Soit un alphabet X avec $a, b, c \in X$ et α une exp. régulière sur X.

Notation	Equivalence	Signification
$[abc]$	$a + b + c$	un symbole parmi un ens. de symboles
$[\hat{}abc]$	$X \backslash \{a, b, c\}$	tous les symboles de X sauf ...
$[a - z]$	$a + b + \cdots + y + z$	n'importe quelle lettre
$[0 - 9]$	$0 + \cdots + 9$	n'importe quel chiffre
$\alpha^?$	$\epsilon + \alpha$	0 ou une fois α
α^+	$\alpha\alpha^*$	au moins une fois α
$\alpha\{n\}$	α^n	α répétée exactement n fois
$\alpha\{n,\}$	$\alpha^n\alpha^*$	α répétée au moins n fois
$\alpha\{,m\}$	$\epsilon + \alpha + \alpha^2 + \cdots + \alpha^m$	α répétée au plus m fois
$\alpha\{n, m\}$	$\alpha^n + \alpha^{n+1} + \cdots + \alpha^m$	α répétée entre n et m fois ($n \leq m$)
$\hat{}\alpha$	αX^*	α doit apparaître en début de ligne
$\alpha\$$	$X^*\alpha$	α doit apparaître en fin de ligne

Pour vous entraîner : *grep -E*, *sed*, *awk*, ...

Expressions régulières

Langage représenté par une expression régulière

Soit r une expression régulière. $\mathcal{L}(r)$ est le **langage représenté par** r.

1. $\mathcal{L}(\emptyset) = \emptyset$, $\mathcal{L}(\epsilon) = \{\epsilon\}$
2. $\forall a \in \Sigma$, $\mathcal{L}(a) = \{a\}$
3. $\mathcal{L}(\alpha, \beta) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta) = \mathcal{L}(\alpha) + \mathcal{L}(\beta)$
4. $\mathcal{L}(\alpha.\beta) = \mathcal{L}(\alpha).\mathcal{L}(\beta)$
5. $\mathcal{L}((\alpha)^*) = (\mathcal{L}(\alpha))^*$

Expressions régulières

Langage représenté par une expression régulière

Soit r une expression régulière. $\mathcal{L}(r)$ est le **langage représenté par** r.

1. $\mathcal{L}(\emptyset) = \emptyset$, $\mathcal{L}(\epsilon) = \{\epsilon\}$
2. $\forall a \in \Sigma$, $\mathcal{L}(a) = \{a\}$
3. $\mathcal{L}(\alpha, \beta) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta) = \mathcal{L}(\alpha) + \mathcal{L}(\beta)$
4. $\mathcal{L}(\alpha.\beta) = \mathcal{L}(\alpha).\mathcal{L}(\beta)$
5. $\mathcal{L}((\alpha)^*) = (\mathcal{L}(\alpha))^*$

Par exemple :

- $\mathcal{L}(a^*) = \{\epsilon, a, aa, aaa, aaaa, \dots\}$
- $\mathcal{L}(c(ab)^*) = \{c, cab, cabab, cababab, \dots\}$
- $\mathcal{L}((a+b)^*) = \{\epsilon, a, aa, b, bb, ab, ba, abba, \dots\}$

Expressions régulières

Langage représenté par une expression régulière

Soit r une expression régulière. $\mathcal{L}(r)$ est le **langage représenté par** r.

1. $\mathcal{L}(\emptyset) = \emptyset$, $\mathcal{L}(\epsilon) = \{\epsilon\}$
2. $\forall a \in \Sigma$, $\mathcal{L}(a) = \{a\}$
3. $\mathcal{L}(\alpha, \beta) = \mathcal{L}(\alpha) \cup \mathcal{L}(\beta) = \mathcal{L}(\alpha) + \mathcal{L}(\beta)$
4. $\mathcal{L}(\alpha.\beta) = \mathcal{L}(\alpha).\mathcal{L}(\beta)$
5. $\mathcal{L}((\alpha)^*) = (\mathcal{L}(\alpha))^*$

Par exemple :

- $\mathcal{L}(a^*) = \{\epsilon, a, aa, aaa, aaaa, \ldots\}$
- $\mathcal{L}(c(ab)^*) = \{c, cab, cabab, cababab, \ldots\}$
- $\mathcal{L}((a+b)^*) = \{\epsilon, a, aa, b, bb, ab, ba, abba, \ldots\}$

Lorsqu'il n'y a pas de confusion possible, on omet le $\mathcal{L}$, et on associe directement l'expression régulière avec le langage associé.

Exemples d'expressions régulières utiles

Numéro de téléphone en France : $\hat{}0[1-9][0-9]\{8\}\$$

Adresse email : $\hat{}[a-zA-Z-]^{+}@[a-zA-Z-]^{+}.[a-zA-Z]\{2,6\}\$$

Nombre : $\hat{}[+-]^{?}[0-9]^{+}(,[0-9]^{*})^{?}([eE][+-]^{?}[0-9]^{+})^{?}\$$

Identifiant : $\hat{}[a-zA-Z\_][a-zA-Z0-9\_]^{*}\$$