Chapitre 1 : Langages réguliers et automates (1)

1 Mots et langages, rappels

Définition 1.1

Un **alphabet** est un ensemble fini de symboles, aussi appelés **caractères** ou **lettres**.

Exemple 1.2

On utilisera souvent les lettres de l'alphabet latin 'a', 'b',

Définition 1.3

Soit Σ un alphabet non vide.

Un **mot** sur Σ est une suite finie de caractères de Σ .

La **longueur** d'un mot x , notée $|x|$, est le nombre de caractères dans x comptés avec multiplicité.

Si $|x|=n \in \mathbb{N}^*$, on indexe les caractères de x par $\llbracket 1, n \rrbracket$, le mot x s'écrit alors $x_1 x_2 \dots x_n$.

Définition 1.4

Peu importe l'alphabet, il existe un seul mot de longueur nulle : **le mot vide**. On le note ε .

Remarque 1.5

Le mot vide ε n'est donc **pas** une lettre.

Définition 1.6

Soit Σ un alphabet non vide.

On note Σ^* l'ensemble des mots sur Σ , et Σ^+ l'ensemble des mots non vides sur Σ .

De plus pour $n \in \mathbb{N}$, on note Σ^n l'ensemble des mots sur Σ de longueur n .

■ Exercice de cours 1.7

Donner une définition de Σ^+ (resp. de Σ^*) comme une union de Σ^n .

Définition 1.8

Soit Σ un alphabet non vide. Soient x et y deux mots sur Σ de longueur respectives m et n .

La **concaténation** des mots x et y , notée $x \cdot y$ et y est le mot $x_1 x_2 \dots x_m y_1 y_2 \dots y_n$.

Remarque 1.9

Intermède mathématique. On appelle aussi concaténation l'opération qui à x et y associe $x \cdot y$.

- Puisque cette opération envoie deux mots de Σ^* sur un mot de Σ^* , on parle d'opération interne, (en faisant cette opération on reste à l'intérieur de l'ensemble). Formellement : $\forall (x, y) \in (\Sigma^*)^2, (x, y) \in \Sigma^*$.
- Cette opération admet un élément neutre : le mot vide. Cela signifie qu'en concaténant un mot avec le mot vide, avant ou après, on ne change pas le mot. Formellement : $\forall x \in \Sigma^*, \varepsilon \cdot x = x \cdot \varepsilon = x$.
- Cette opération est associative, cela signifie que concaténer y à x , puis z au mot obtenu, revient au même que d'abord concaténer z à y , puis le mot obtenu à x . Formellement : $\forall (x, y, z) \in (\Sigma^*)^3, (x \cdot y) \cdot z = x \cdot (y \cdot z)$.

En algèbre, on dit que $(\Sigma^*, \cdot)$ est un monoïde pour désigner cette structure.

Remarque 1.10

Notez que l'opération de concaténation n'est pas commutative dès que l'alphabet Σ a plus de deux lettres, c'est-à-dire qu'en général $x \cdot y$ et $y \cdot x$ ne sont pas les mêmes mots.

📌 Exercice de cours 1.11

Donner un exemple illustrant la remarque précédente.

Définition 1.12

Soient x et y deux mots sur un alphabet non vide Σ . Notons $n = |x|$ et $m = |y|$.

- On dit que y est un **facteur** de x dès lors qu'il existe deux mots u et v sur Σ tels que $x = u \cdot y \cdot v$.
- On dit que y est un **préfixe** de x dès lors qu'il existe un mot v sur Σ tel que $x = y \cdot v$.
- On dit que y est un **suffixe** de x dès lors qu'il existe un mot u sur Σ tel que $x = u \cdot y$.

Remarque 1.13

- Notez que les mots u et v dans la décomposition de x autour du facteur y ne sont pas uniques, tandis que les mots u et v dans la décomposition de x selon le préfixe ou suffixe y le sont.
- Par définition, les préfixes et suffixes sont des facteurs.
- Le mot vide est préfixe et suffixe de n'importe quel mot.
- N'importe quel mot est toujours son propre préfixe et son propre suffixe.

📌 Exercice de cours 1.14

Illustrer et/ou démontrer chacun des points de la remarque ci-dessus en revenant aux définitions.

Définition 1.15

Soient x et y deux mots sur un alphabet non vide Σ . Notons $n = |x|$ et $m = |y|$.

On dit que y est un **sous-mot** de x dès lors qu'il existe σ une fonction strictement croissante de $\llbracket 1, m \rrbracket$ dans $\llbracket 1, n \rrbracket$ telle que $y = x_{\sigma(1)} x_{\sigma(2)} \dots x_{\sigma(m)}$ ♣.

Remarque 1.16

- Par définition, les facteurs sont des sous-mots, mais le contraire est faux.
- Le mot vide est sous-mot de n'importe quel mot.
- N'importe quel mot est toujours sous-mot de lui-même.

📌 Exercice de cours 1.17

Illustrer et/ou démontrer chacun des points de la remarque ci-dessus en revenant aux définitions.

♣. En mathématiques, on dirait que y est une suite extraite de x , à ceci près qu'il s'agit de suites finies.

Exemple 1.18

L'ensemble $\Sigma = \{a, b, c\}$ est un alphabet. $x = ba$, $y = aaa$, $z = babaaa$ sont des mots sur Σ . On a $|x| = 2$, $|y| = 3$ et $|z| = 6$. La concaténation de x et y est $z' = baaaa$, qui est un sous-mot de z , mais pas un facteur de z . Le mot x est un préfixe de z et z' , mais pas de y , qui lui est suffixe de z et z' .

Définition 1.19

Un **langage** sur un alphabet non vide Σ est un ensemble de mots sur Σ , soit une partie de Σ^* .

Exemple 1.20

Supposons dans cet exemple que $\Sigma = \{a, b\}$. Alors les ensembles suivants sont des langages sur Σ : $\emptyset$, Σ^* , $\{abb, aa, baa\}$, $\{\underbrace{aa \dots a}_n \underbrace{bb \dots b}_n \mid n \in \mathbb{N}\}$.

Exercice de cours 1.21

1. Parmi les langages ci-dessus, lesquels sont finis, lesquels sont infinis ?
2. Justifier que si Σ n'est pas vide, l'ensemble des langages sur Σ est infini.

Dans toute la suite du chapitre Σ désignera un alphabet non vide.

2 Langages réguliers

2.1 Opérations sur les langages

Les langages étant définis comme des ensembles, on peut appliquer aux langages les opérations ensemblistes usuelles : l'union, l'intersection, le complémentaire (par rapport à Σ^*), la différence, la différence symétrique. On définit ci-après quelques autres opérations, plus spécifiques aux langages.

Définition 2.1

Soient L_1 et L_2 deux langages sur Σ .

La **concaténation** de L_1 et L_2 est $L_1 \cdot L_2 = \{u \cdot v \mid u \in L_1, v \in L_2\}$.

Exercice de cours 2.2

Donner $\{aa, aaa\} \cdot \{b, bb\}$, $\{aa, aaa\} \cdot \{a, aa\}$ et finalement $\{\underbrace{aa \dots a}_n \mid n \in \mathbb{N}\} \cdot \{\underbrace{bb \dots b}_n \mid n \in \mathbb{N}\}$.

Remarque 2.3

L'opération de concaténation des langages, comme celle sur les mots, est associative et non commutative. L'élément neutre de cette opération est le langage réduit au mot vide : $\{\varepsilon\}$. L'ensemble vide est quant à lui absorbant pour la concaténation. L'opération $\cdot$ est distributive à gauche et à droite par rapport à $\cup$, i.e. $\forall (K, L, M) \in (\mathcal{P}(\Sigma^*))^3$, $K \cdot (L \cup M) = (K \cdot L) \cup (K \cdot M)$ et $(K \cup L) \cdot M = (K \cdot M) \cup (L \cdot M)$.

Définition 2.4

On note en exposant l'itéré de la concaténation, ainsi pour L un langage sur Σ , $L^0 = \{\varepsilon\}$ et pour $n \in \mathbb{N}^*$, $L^n = L \cdot L^{n-1}$. De plus on note $L^* = \bigcup_{n \in \mathbb{N}} L^n$ et $L^+ = \bigcup_{n \in \mathbb{N}^*} L^n$.

L'opération qui à L associe L^* est appelée **étoile de Kleene**.

Exercice de cours 2.5

Dans le cas où $L = \{\varepsilon, aa, b\}$ décrire au moyen d'une phrase en français le langage L^* .

Définition 2.6

Les **opérations régulières** sur les langages sont l'union, la concaténation et l'étoile de Kleene.

Définition 2.7

L'ensemble des **langages réguliers** sur Σ est le plus petit ensemble de $\mathcal{P}(\mathcal{P}(\Sigma^*))$ contenant $\emptyset$ et $\{a\}$ pour tout $a \in \Sigma$ et stable par les opérations régulières.

Remarque 2.8

La définition 2.7 est une définition inductive de l'ensemble des langages réguliers.

Exercice de cours 2.9

Justifier que le langage $\{aa, aab, \varepsilon\}$ est bien un langage régulier.

Justifier que le langage $\{bb, bab, baab, \dots\}$, i.e. $\{b \underbrace{aaa \dots a}_n b, \dots \mid n \in \mathbb{N}\}$, est lui aussi régulier.

Remarque 2.10

Attention, la définition précédente ne dit pas qu'un langage régulier est stable par concaténation, union $\clubsuit$ et passage à l'étoile, elle dit que l'ensemble des langages réguliers est stable par

Exercice de cours 2.11

Donner un langage régulier qui n'est pas stable par concaténation.

2.2 Expressions régulières

Définition 2.12

L'ensemble des **expressions régulières** sur Σ , noté ici $\mathcal{E}_{reg}(\Sigma)$, est défini par induction à partir des règles suivantes :

- le symbole $\emptyset$ est une expression régulière ;
- le symbole ε est une expression régulière ;
- si $a \in \Sigma$ alors a est une expression régulière ;
- si e_1 et e_2 sont deux expressions régulières alors l'expression $e_1 \cdot e_2$ est une expression régulière ;
- si e_1 et e_2 sont deux expressions régulières alors l'expression $e_1 | e_2$ est une expression régulière.
- si e est une expression régulière alors l'expression e^* est une expression régulière.

Exercice de cours 2.13

Donner l'arbre de syntaxe de l'expression régulière $a \cdot (b | (a \cdot \emptyset)^*)$.

Remarque 2.14

Les expressions régulières sont des objets syntaxiques, leur écriture les définit complètement. Ainsi l'expression $a|b$ n'est pas la même que l'expression $b|a$, l'expression $(a.a).b$ pas la même que $a.(a.b)$ (Mais la sémantique définie ci-dessous apporte une notion d'équivalence qui règle tout ça).

Définition 2.15

Le langage associé à une expression régulière sur Σ est donné par la fonction $\mathcal{L}$ définie inductivement comme suit.

$$\mathcal{L} = \left(\begin{array}{ll} \mathcal{E}_{reg}(\Sigma) & \rightarrow \mathcal{P}(\Sigma^*) \\ \emptyset & \mapsto \emptyset \\ \varepsilon & \mapsto \{\varepsilon\} \\ a \in \Sigma & \mapsto \{a\} \\ E^* & \mapsto \mathcal{L}(E)^* \\ E_1|E_2 & \mapsto \mathcal{L}(E_1) \cup \mathcal{L}(E_2) \\ E_1 \cdot E_2 & \mapsto \mathcal{L}(E_1) \cdot \mathcal{L}(E_2) \end{array} \right)$$

Exercice de cours 2.16

Donner le langage de l'expression régulière $(b \cdot b \cdot b)^*(a|c|\varepsilon)$.

Proposition 2.17

L'ensemble des langages réguliers est exactement l'image de $\mathcal{L}$.

Remarque 2.18

La structure inductive de $\mathcal{E}_{reg}(\Sigma)$ permet de définir de manière inductive des fonctions sur $\mathcal{E}_{reg}(\Sigma)$. Ainsi, on peut définir comme suit la fonction qui associe à une expression régulière l'ensemble des lettres qui y apparaissent.

$$\text{lettres} = \left(\begin{array}{ll} \mathcal{E}_{reg}(\Sigma) & \rightarrow \mathcal{P}(\Sigma) \\ \emptyset & \mapsto \emptyset \\ \varepsilon & \mapsto \emptyset \\ a \in \Sigma & \mapsto \{a\} \\ E^* & \mapsto \text{lettres}(E) \\ E_1|E_2 & \mapsto \text{lettres}(E_1) \cup \text{lettres}(E_2) \\ E_1 \cdot E_2 & \mapsto \text{lettres}(E_1) \cup \text{lettres}(E_2) \end{array} \right)$$

Exercice de cours 2.19

Donner $\text{lettres}((b \cdot b \cdot b)^*(a|c|\varepsilon))$.

3 Automates finis (sans ε -transition)

3.1 Définitions

Définition 3.1

Un **automate fini (sans ε -transition)** est un quintuplet $\mathcal{A} = (\Sigma, Q, I, F, \delta)$ où :

- Σ est un alphabet;
- $I \subseteq Q$;
- $\delta \subseteq Q \times \Sigma \times Q$.
- Q est un ensemble fini;
- $F \subseteq Q$;

Les éléments de Q (resp. I, F) sont appelés **états** (resp. **état initiaux**, **état finaux**).

De plus si $t = (q, a, q') \in \delta$, on dit que t est une **transition** de l'état q à l'état q' d'étiquette a , ce que l'on note parfois $q \xrightarrow{a} q'$.