INF 302 : Langages & Automates

Chapitre 2 : Notions préliminaires - alphabet, mot, langage

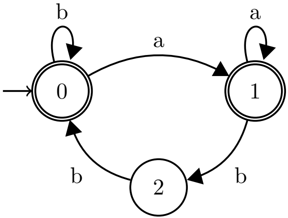
Yliès Falcone

ylies.falcone@univ-grenoble-alpes.fr — www.ylies.fr

Univ. Grenoble-Alpes, Inria
-
Laboratoire d'Informatique de Grenoble - www.liglab.fr
Équipe de recherche LIG-Inria, Corse - team.inria.fr/corse/

Année Académique 2020 - 2021

Intuition et objectifs



Définir (mathématiquement) quelques ingrédients de base des automates :

- symbole, alphabet, mot — *syntaxe*.
- mot, langage, quelques opérations de composition de langages – *sémantique*.

Symboles et alphabets

Un automate lit des symboles.

L'ensemble des symboles lus par un automate est appelé son alphabet.

Définition (Alphabet)

Un **alphabet** est un ensemble fini dont les éléments sont appelés **symboles**.

Exemple (Alphabets)

- Pour l'exemple de la transaction électronique (du chapitre précédent), l'alphabet est $\{abd, pay, send, sol, tra\}$.
- $\{a, b, c, \ldots, z\}$ peut être l'alphabet d'un automate utilisé pour reconnaître les mots d'une langue latine.
- $\{0, 1\}$ est l'alphabet des nombres représentés en notation binaire.
- L'ensemble des caractères ASCII.

L'alphabet d'un automate est usuellement noté Σ.

Mots

Vision application

Intuitivement, un **mot** sur un alphabet Σ est une séquence finie (possiblement vide) de symboles dans Σ.

Définition (Mot)

Un **mot de longueur** $n \in \mathbb{N}$ est une *application* de $[1, n]$ vers Σ.

Remarque C'est une application, et non pas une fonction quelconque. □

Longueur d'un mot

Nombre de symboles dans ce mot.
On note $|u|$ la *longueur* du mot u.

Définition (Mot vide)

- Le mot vide est la fonction de l'ensemble vide ($\emptyset$) vers Σ.
- Le mot vide est noté ϵ_Σ ou ϵ, lorsque le contexte est clair.

Exemple (Mot)

Considérons l'alphabet $\Sigma = \{a, b\}$:

- ϵ est le mot de longueur 0,
- a et b sont les mots de longueur 1,
- aa, ab, ba et bb sont les mots de longueur 2.

Exemple (Symbole à une position)

- $ab(1) = a$.
- $ab(2) = b$.
- $ab(3)$ n'est pas défini.

Mots

Vision inductive

De façon équivalente, nous pouvons définir les mots de manière inductive.

Définition (Mot (construit par la droite))

Considérons un alphabet Σ.

- ϵ est le mot de longueur 0 sur Σ,
- si u est un mot de longueur $n \in \mathbb{N}$ sur Σ et a un symbole de Σ, alors ua est un mot sur Σ de longueur $n + 1$.

Remarque

- Cette définition est équivalente à la précédente. Intuition : les applications de $\{1, n\}$ vers Σ définissent les séquences de longueur n sur Σ.
- Nous pouvons définir également les mots *construits par la gauche* de façon similaire. (Les deux définitions sont équivalentes.)
- Nous utiliserons la définition la plus pratique en fonction de la situation.

Langages

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

Définition (Langage)

Un langage sur l'alphabet Σ est un ensemble de mots sur Σ ;

- c-à-d un sous-ensemble de Σ^* ;
- c-à-d un élément de l'ensemble $\mathcal{P}(\Sigma^*)$, où $\mathcal{P}(\Sigma^*)$ dénote l'ensemble des sous-ensembles de Σ^*.

Exemple (Langage)

- $\emptyset$ est un langage sur Σ : langage vide,
- Σ^* est un langage sur Σ : langage universel,
- $\{\epsilon\}$ est un langage sur Σ,
- $\{0, 00, 001\}$ est aussi un langage sur l'alphabet $\{0, 1\}$,
- L'ensemble des mots qui contiennent un nombre impair de 0 est un langage sur n'importe quel alphabet Σ tel que $\Sigma \supseteq \{0\}$,
- L'ensemble des mots qui contiennent autant de 0 que de 1 est un langage sur n'importe quel alphabet Σ tel que $\Sigma \supseteq \{0, 1\}$.

Concaténation de mots

Intuitivement :

- la concaténation des mots 01 et 10 est le mot 0110,
- la concaténation du mot vide ϵ et du mot 101 est le mot 101.

$\hookrightarrow$ la concaténation de deux mots (applications) est un mot (une application).

◁ Concaténation de deux mots

Considérons un alphabet Σ.

Définition (Concatenation)

- La concaténation est une application de $\Sigma^* \times \Sigma^*$ vers Σ^*,
- La concaténation de deux mots u et v dans Σ^* est le mot $u \cdot v : [1, |u| + |v|] \to \Sigma$ tel que :

$$(u \cdot v)(i) \stackrel{\text{def}}{=} \begin{cases} u(i) & \text{si } i \in [1, |u|] \\ v(i - |u|) & \text{si } i \in [|u| + 1, |u| + |v|] \end{cases}$$

ϵ est élément neutre à droite et à gauche de la concaténation.

$\forall u \in \Sigma^* : u \cdot \epsilon_\Sigma = \epsilon_\Sigma \cdot u = u.$

Remarque L'opérateur de concaténation pourra être omis pour des raisons de lisibilité et nous noterons, par exemple, uv au lieu de $u \cdot v$.

Préfixes, suffixes et facteurs d'un mot

Des « sous-mots » particuliers

Définition (Préfixe, suffixe et facteur)

Considérons deux mots u et v sur un alphabet Σ.

- v est un préfixe de u, noté $v \preceq u$, s'il existe un mot v' (sur Σ) tel que $v \cdot v' = u$.
- v est un suffixe de u, s'il existe un mot v' (sur Σ) tel que $v' \cdot v = u$.
- v est un facteur de u, s'il existe deux mots v' et v'' (sur Σ) tels que $v' \cdot v \cdot v'' = u$.
- v est une extension de u, si u est un préfixe de v.

Remarque

- Les préfixes et les suffixes d'un mot en sont des facteurs particuliers.
- Un mot est préfixe, suffixe et facteur de lui même.
- Le mot vide est préfixe, suffixe et facteur de tout mot.
- (Préfixes/suffixes sont aussi appelés facteurs gauches/droits.)

Exemple (Préfixes, suffixes et facteurs du mot $abccba$ sur l'alphabet $\{a, b, c\}$)

- préfixes : ϵ, a, ab, abc, $abcc$, $abccb$, $abccba$.
- suffixes : ϵ, a, ba, cba, $ccba$, $bccba$, $abccba$.
- $bccb$, bcc, bc, cb, cc, ccb, c sont des facteurs (en plus des préfixes et suffixes).

Concaténation de langages

Extension de la concaténation des mots aux langages

Intuition : À partir de deux langages, la concaténation de langages produit un nouveau langage contenant tous les mots construits en concaténant un mot du premier langage avec un mot du deuxième.

◁ Concaténation de deux langages

Définition (Concaténation de langages)

$$\begin{array}{rcl} \cdot & : & \mathcal{P}(\Sigma^*) \times \mathcal{P}(\Sigma^*) \to \mathcal{P}(\Sigma^*) \\ L_1 \cdot L_2 & \stackrel{\text{def}}{=} & \{u_1 \cdot u_2 \mid u_1 \in L_1 \wedge u_2 \in L_2\} \end{array}$$

Remarque Pas de commutativité :

$$L_1 \cdot L_2 \neq L_2 \cdot L_1 \quad \text{(en général)}$$

□

Exemple (Concaténation de langages)

Considérons les langages $\{a, aa\}$ et $\{b, bb\}$.

$$\{a, aa\} \cdot \{b, bb\} = \{ab, abb, aab, aabb\}$$

Exemple (Concaténation de langages)

Soit un langage L sur un alphabet Σ :

- $L \cdot \emptyset = \emptyset$ $(= \{u_1 \cdot u_2 \mid u_1 \in L \wedge u_2 \in \emptyset\})$,
- si $L \neq \Sigma^*$ et $\epsilon \notin L$, alors $L \cdot \Sigma^* \subset \Sigma^*$,
- Si $\epsilon \in L$ alors $L \cdot \Sigma^* = \Sigma^*$,
- $L \cdot \{\epsilon\} = \{\epsilon\} \cdot L = L$
  $(= \{u_1 \cdot u_2 \mid u_1 \in L \wedge u_2 \in \{\epsilon\}\})$.

Fermeture par préfixe, suffixe et extension

Définition

Soit L un langage sur un alphabet Σ.

Définition (Fermeture par préfixe et par suffixe)

- La fermeture par préfixe de L, noté $\mathrm{Pref}(L)$, est le langage formé par *l'ensemble préfixes des mots de L*, défini comme :

  $$\mathrm{Pref}(L) = \left\{ w \in \Sigma^* \mid \exists w' \in \Sigma^* : w \cdot w' \in L \right\}$$

  de manière équivalente : $\mathrm{Pref}(L) = \{ w \in \Sigma^* \mid \exists w' \in L : w \preceq w' \}$.
- La fermeture par suffixe de L, noté $\mathrm{Suf}(L)$, est le langage formé par *l'ensemble des suffixes des mots de L*, défini comme :

  $$\mathrm{Suf}(L) = \left\{ w \in \Sigma^* \mid \exists w' \in \Sigma^* : w' \cdot w \in L \right\}$$

- La fermeture par extension de L, noté $\mathrm{Ext}(L)$, est le langage formé par *l'ensemble des extensions des mots de L*, défini comme :

  $$\mathrm{Ext}(L) = \left\{ w \in \Sigma^* \mid \exists w' \in L : w' \preceq w \right\}$$

On a : $L \subseteq \mathrm{Pref}(L)$, $L \subseteq \mathrm{Suf}(L)$ et $L \subseteq \mathrm{Ext}(L)$.

Fermeture par préfixe, suffixe et extension

Exemples

Exemple (Fermeture par préfixe et par suffixe)

Pour $L = \{abcd, xyz\}$:

- $\mathrm{Pref}(L) = \{\epsilon, a, ab, abc, abcd, x, xy, xyz\}$;
- $\mathrm{Suf}(L) = \{\epsilon, d, cd, bcd, abcd, z, yz, xyz\}$;
- $\mathrm{Ext}(L)$ est l'ensemble des mots commençant par *abcd* ou *xyz*.

Exemple (Fermeture par préfixe et par suffixe)

Pour L défini comme l'ensemble des mots formés par répétitions (possiblement 0 fois) du mot $a \cdot b$:

- $\mathrm{Pref}(L)$ est l'ensemble des mots formés par répétitions (possiblement 0 fois) du mot $a \cdot b$ et terminant possiblement par a après la dernière répétition de $a \cdot b$.
- $\mathrm{Suf}(L)$ est l'ensemble des mots commençant par b ou ϵ et suivi d'une répétition (possiblement 0 fois) du mot $a \cdot b$.
- $\mathrm{Ext}(L)$ est l'ensemble des mots commençant par une répétition (possiblement 0 fois) du mot $a \cdot b$.

Langages fermés par préfixe, suffixe et extension

Définition (Langage fermé par préfixe et par suffixe)

Un langage L sur un alphabet donné est dit :

- fermé par préfixe (de manière équivalente préfixe-clos) si $L = \mathrm{Pref}(L)$;
- fermé par suffixe (de manière équivalente suffixe-clos) si $L = \mathrm{Suf}(L)$.
- fermé par extension (de manière équivalente extension-clos) si $L = \mathrm{Ext}(L)$.

Exemple (Fermeture par préfixe et par suffixe)

Sur tout alphabet Σ, le langage universel (Σ^*) est fermé par préfixe, suffixe et extension.

Remarque Dans les chapitres suivants, nous verrons des notations pour décrire de manière plus concise et rigoureuse les langages. Nous reviendrons sur ces propriétés de fermeture. □

Fermeture de Kleene d'un langage

Intuition : La fermeture de Kleene d'un langage L est le langage (c-à-d l'ensemble des mots) formés par des mots pris dans L.

Définition (Fermeture de Kleene)

La fermeture de Kleene d'un langage L, notée L^*, est l'ensemble défini inductivement par les deux règles suivantes :

- $\epsilon \in L^*$ et
- si $u \in L$, $v \in L^*$, alors $u \cdot v \in L^*$,
- (de manière équivalente à la précédente règle, si $u \in L^*, v \in L$, alors $u \cdot v \in L^*$).

Remarque La fermeture de Kleene de L est le langage des mots formés par un nombre fini de concaténations de mots de L :

$$L^* = \{\epsilon\} \cup \{u_0 \cdots u_n \mid n \in \mathbb{N} \wedge (\forall i : i \leq n \implies u_i \in L)\}$$

Exemple (Fermeture de Kleene)

Pour $L = \{ab, cd\}$, $L^* = \{\epsilon, ab, cd, abab, abcd, cdcd, cdab, ababab, \ldots\}$.

Résumé du Chapitre 1 : notions préliminaires pour les automates

- **alphabet** : ensemble de symboles (auxquels un automate réagit),
- **mot** : séquence de symbole (entrée d'un automate),
- **préfixe**, **suffixe**, **facteur** et **extension** d'un mot :

```
| u n   m o t |
|préfixe->|
           |<-suffixe|
    |<-facteur->|
| ...e..x..t..e..n..s|..i..o..n..->|
```

- **langage** : ensemble de mots,
- **concaténation de mots** : mettre un mot à la suite d'un autre pour former un mot,
- **concaténation de langages** : tous les mots formés en concaténant les mots du premier langage aux mots du second ;
- **fermeture par préfixe, suffixe et extension** d'un langage : langage contenant tous les préfixes/suffixes/extensions des mots d'un langage ;
- langage **préfixe-clos**, langage **suffixe-clos**, langage **extension-clos** : un langage qui contient tous ses préfixes/suffixes/extensions ;
- **fermeture de Kleene** d'un langage : langage des mots formés par concaténations de mots du langage considéré.