Institut International de Technologie
Cours Théorie des langages & automates

Introduction aux langages

INTERNATIONAL INSTITUTE OF TECHNOLOGY
North American Private University

Soufiene.Lajmi@gmail.com

Année universitaire 2020-2021

Plan

(I) Introduction et concepts fondamentaux
(II) Les langages réguliers et les expressions régulières
(III) Les grammaires
(IV) Les automates

Evaluation= 30%CC + 70% Examen final

Chapitre 1
Introduction et concepts fondamentaux

Sommaire

Alphabet, mots et langage

Opérations sur les mots

Structure de monoïde

Opérations sur les langages

Alphabet, mots et langage

- Définition d'un alphabet :
  - Un **alphabet** est un ensemble fini d'éléments appelés **symboles** (ou **lettres**).
  - Exemples de symboles et d'alphabet :
    - Lettres (minuscules et majuscules) : a, b, ... z, A, B, ..., Z.
    - Chiffres décimaux : 0, 1, ..., 9
    - Les caractères du clavier.
    - Alphabet des nombres binaires : {0, 1}.
    - Alphabet informatique : ASCII et EBCDIC.

Alphabet, mots et langage

- Exemples de symboles et d'alphabet (suite) :
  - Alphabet des nombres octaux : {0, 1, ..., 7}.
  - Alphabet des nombres hexadécimaux : {0, 1, ..., 9, a, b, c, d, e, f, A, B, C, D, E, F}.
  - Alphabet Français : {a, A, b, B, ..., z, Z}.
  - Alphabet des identificateurs du langage C : {a, b, ..., z, A, B, ..., Z, 0, 1, ..., 9, _}.
  - Alphabet des nombres entiers décimaux en C : {+, -, 0, 1, ..., 9}.
- Remarque : l'alphabet vide ne contient aucun symbole.

Alphabet, mots et langage

- Définition d'un mot sur un alphabet :
  - Un **mot** s sur un alphabet A est une séquence finie de symboles de A.
  - Les symboles faisant partie d'un mot s s'écrivent juxtaposés l'un à l'autre de gauche à droite.
  - La **longueur d'un mot** s sur l'alphabet A est le nombre des symboles constituants le mot s. Elle est notée indifféremment par **|s|** ou **lg(s)**.
  - Sur tout alphabet A, on peut définir un mot spécial : le mot vide noté **ε**. sa longueur est 0 ($|\varepsilon| = 0$). On a : $|s| = 0$, si et seulement si $s = \varepsilon$.

Alphabet, mots et langage

- Exemples de mots sur des alphabets :
  - s = aller est un mot sur l'alphabet Français. |s| = ?.
  - s = 001 est un mot sur l'alphabet {0, 1}. |s| = ?.
  - s = 11AF0009C est un mot sur l'alphabet des nombres hexadécimaux. |s| = ?.
  - s = -12.254 est un mot sur l'alphabet des nombres. On a : |s| = ?.
- Soit s un mot de longueur n sur un alphabet A. Si n = 0, alors s est le mot vide ($s = \varepsilon$). Si n > 0, s s'écrit sous la forme $\alpha_1\alpha_2\ldots\alpha_n$ où chaque α_i est un symbole ($\alpha_i \in A$) pour tout indice *i* de 1 à n. On a : $s = \alpha_1\alpha_2\ldots\alpha_n$.

Alphabet, mots et langage

- Dans un mot $s = \alpha_1\alpha_2\ldots\alpha_n$ de longueur n sur un alphabet A, α_i s'appelle le **symbole** (ou **terme**) **de rang** i.
- Le symbole de rang i du mot s, se note par s(i).
- Parfois, s(i) se nomme aussi la « i ^{ème} lettre de s ».
- Deux mots $s = \alpha_1\alpha_2\ldots\alpha_n$ et $t = \beta_1\beta_2\ldots\beta_m$ sur un même alphabet A sont égaux, si et seulement si :
  - $|s| = |t|$ (c'est-à-dire ici, n = m).
  - Pour tout i de 1 à n, s(i) = t(i) (c'est-à-dire ici $\alpha_i = \beta_i$).
- Deux mots $s = \alpha_1\alpha_2\ldots\alpha_n$ et $t = \beta_1\beta_2\ldots\beta_m$ sur un même alphabet A sont différents, si et seulement si $|s| \neq |t|$ ou il existe un indice i tel que : $s(i) \neq t(i)$.

Alphabet, mots et langage

- Définition d'un langage sur un alphabet :
  - Un **langage** L <u>sur un alphabet A</u> est un <u>ensemble de mots</u> sur A.
- Exemples :
  - L = {0, 10, 101} est un langage sur l'alphabet {0, 1}.
  - V = {1, 01, 11, 001, 011, 101, 111, 0001, 0011, 0101, 0111, …} est le langage de tous les mots sur {0, 1} se terminant par un « 1 ».
- Remarque :
  - Un langage sur un alphabet A peut être vide, fini ou infini.

Alphabet, mots et langage

- Certains langages particuliers sur un alphabet A donné :
  - Un alphabet A est un langage sur A. En fait, un symbole est considéré comme un mot de longueur 1.
  - L'ensemble vide $\varnothing$ est un langage sur A.
  - $\{\varepsilon\}$ l'ensemble formé par le mot vide est aussi un langage sur A. Attention, $\{\varepsilon\} \neq \varnothing$.
  - L'ensemble formé par tous les mots possibles sur l'alphabet A est un langage sur A. Ce langage très important est noté par A^*.
  - A^* s'appelle le **monoïde libre** sur A.

Alphabet, mots et langage

- Exemples du monoïde libre A^* sur A :
  - Si $A = \varnothing$ (alphabet vide), alors $A^* = \{\varepsilon\}$.
  - Si $A = \{\varepsilon\}$ (alphabet formé par le symbole vide), alors $A^* = \{\varepsilon\}$.
  - Si $A = \{a\}$ (alphabet à un seul symbole), alors $A^* = ?$
  - Si $A = \{0, 1\}$ (alphabet binaire), alors $A^* = ?$.
- Propriétés de base du langage A^* :
  - A^* n'est jamais vide. En effet, on a toujours $\varepsilon \in A^*$.
  - $A \subset A^*$.

Alphabet, mots et langage

- $A^* = \{\varepsilon\}$ si et seulement si $A = \emptyset$ ou $A = \{\varepsilon\}$.
- A^* est infini dès que A contient au moins un symbole (qui ne soit pas le symbole vide).
- Tout langage L sur un alphabet A est une partie de l'ensemble de A^* : L est un langage sur l'alphabet A, si et seulement si $L \subset A^*$.
- Un mot s sur l'alphabet A est simplement un élément de l'ensemble A^*.

Alphabet, mots et langage

- Définition d'un langage formel:
  - C'est tout sous ensemble de A* dont les mots peuvent être définis de deux façons conformément à:
  - Définition par propriété: il s'agit d'une modélisation formelle d'une description naturelle d'un langage.
    - Exemple L1 = {ensemble des mots definis sur {a,b} de longueur paire}

      $L1 = \{w \in \{a,b\}^* \ / \ |w| = 2n \text{ avec } n >= 0\}$
  - Définition récursive : Définition dans laquelle, un langage est défini sur lui même.
  - Exemple:
  - $L2 = \{w \in A^* \ / \ w = a \text{ ou } w = aw1 \ ; \ w1 \in L2\}$

    $= \{a, aa, .., aaaa, ...\}$
  - $L3 = \{w \in A^* \ / \ w = \varepsilon \text{ ou } w = w1w2 \text{ avec } |w1| = 2 \text{ et } w2 \in L3\}$
  - $L3 \equiv L1$

Opérations sur les mots

- **Concaténation** (**collage**, **jointure**) :
  - Soit s et t deux mots sur un même alphabet A.
  - La concaténation de s et t est le mot obtenu en écrivant d'abord les symboles de s, suivis ensuite par les symboles de t.
  - Formellement, si $s = \alpha_1\alpha_2\ldots\alpha_n$ et $t = \beta_1\beta_2\ldots\beta_m$ alors, la concaténation de s et de t, notée « st » donne le mot suivant : $st = \delta_1\delta_2\ldots\delta_{n+m}$ où :
    - $\delta_i = \alpha_i$ pour tout indice i de 1 à n.
    - $\delta_i = \beta_{i-n}$ pour tout indice i de n + 1 à n + m.

15

Opérations sur les mots

- Exemple :
  - Soit les mots s = 01 et t = 11 sur A = {0, 1}. Alors, la concaténation de s et t est st = ?.
  - La concaténation de t et s est ts = ?.
- Propriétés de la concaténation :
  - **Associativité** : (st)u = s(tu) = stu.
  - **Élément neutre** : $\varepsilon s = s\varepsilon = s$, pour tout mot s sur A.
  - $|st| = |s| + |t|$.
  - **Non commutative** (en général) : $st \neq ts$.
  - $|st| = 0$ si et seulement si $s = t = \varepsilon$.

16

Opérations sur les mots

- **Puissance** :
  - Soient s un mot sur un alphabet A et n un entier naturel.
  - La puissance n ^{ème} de s, notée s^n, est définie de façon récursive par :
    - $s^0 = \varepsilon$.
    - $s^n = s^{n-1}s$, pour $n \geq 1$
  - Exemple :
    - Soient s = aba un mot sur l'alphabet A = {a, b}. Alors, s^2 = abaaba, s^3 = abaabaaba.

Opérations sur les mots

- **Propriétés de la puissance** :
  - Soit A un alphabet. On montre facilement les propriétés suivantes :
    - $\forall s \in A^* : s^0 = \varepsilon$ et $s^1 = s$.
    - $\forall s \in A^*, \forall n, m \in N, : s^n s^m = s^{n+m}$.
    - $\forall s \in A^*, \forall n, m \in N, : (s^n)^m = s^{nm}$
    - $\forall s \in A^*, \forall n \in N, : |s^n| = n|s|$.

Opérations sur les mots

- **Miroir (renversé)** :
  - Soit A un alphabet et u un mot sur A.
  - Le miroir (ou le renversé) de u est le mot obtenu à partir de u en inversant les lettres de celui-ci.
  - Le miroir d'un mot u est noté par u^R ou ũ.
  - Exemple :
    - le miroir du mot u = abab sur l'alphabet A = {a, b} est ũ = baba.
  - Un mot u sur A est un **palindrome** si et seulement si u est identique (égal) à son miroir ũ (u = ũ).

Opérations sur les mots

- Exemples :
  - 101 est un palindrome sur l'alphabet A = {0, 1}.
  - abcba est un palindrome sur A = {a, b, c}.
  - Quelques palindromes Français :
    - radar.
    - non.
    - alla, ressasser.
    - elle, été.
    - ici.
    - tôt, ses.

Opérations sur les mots

- **Propriétés du miroir** :
  - Soit A un alphabet. On montre facilement les propriétés suivantes :
    - $\forall$ u, v $\in A^*$: $(uv)^R = v^R u^R$.
    - $\forall$ u $\in A^*$: $(u^R)^R = u$ (le miroir est involutif).
    - $\forall$ s $\in A^*$: $|s^R| = |s|$ (le mot et son miroir ont la même longueur).

Opérations sur les mots

- **Préfixes, suffixes, facteur**:
  - Soient A un alphabet et s un mot sur A.
  - On dit qu'un mot u sur A est un **préfixe** de s si et seulement si il existe un mot v sur A tel que s = uv.
  - On dit qu'un mot v sur A est un **suffixe** de s si et seulement si il existe un mot u sur A tel que s = uv.
  - On dit qu'un mot w sur A est un **facteur** de s si et seulement si il existe deux mots u et v sur A tels que s = uwv.

Opérations sur les mots

- Exemples : soit le mot s = abbba sur A = {a, b}.
  - Les préfixes de s sont : ε, a, ab, abb, abbb et abbba.
  - Les suffixes de s sont : ε, a, ba, bba, bbba et abbba.
  - Les facteurs de s sont : ε, a, b, ab, bb, ba, abb, bbb, bba, abbb, bbba et abbba.
- Remarque :
  - ε et s sont des préfixes, des suffixes et des facteurs de s.

Opérations sur les mots

- On appelle **préfixe** (resp. **suffixe** ou **facteur**) **propre** d'un mot s un préfixe (resp. suffixe ou facteur) différent de ε et s.
- Exemples :
  - Les préfixes propres de s sont : a, ab, abb et abbb.
  - Les suffixes propres de s sont : a, ba, bba et bbba.
  - Les facteurs propres de s sont : a, b, ab, bb, ba, abb, bbb,bba, abbb et bbba.

Structure de monoïde

- En algèbre, un **monoïde** est un ensemble M muni d'une loi de composition interne * (c'est-à-dire une application de M×M vers M) qui :
  - soit associative : (x * y) * z = x * (y * z).
  - admet un élément neutre e : x * e = e * x = x.
- Exemples :
  - (N, +, 0) est un monoïde.
  - (N, ×, 1) est un monoïde.
  - Tout groupe (G, *, e) est un monoïde.
  - $(A^*, \underset{\text{concat}}{\cdot}, \varepsilon)$

Structure de monoïde

- **Théorème 1** :
  - L'ensemble de tous les mots sur un alphabet A est un monoïde pour l'opération de la concaténation des mots.
- Cet ensemble s'appelle le **monoïde libre** sur A et nous l'avons noté par A^*.
- On note A^+ l'ensemble des mots non vides sur A : A^+ est donc $A^* \setminus \{\varepsilon\}$.
- Exemple :
  - pour A = {a}, A^+ = {a, aa, aaa, aaaa, …}.
  - pour A = {a, b}, A^+ = {a, b, aa, ab, ba, bb, …}.

Opérations sur les langages

- Soient A un alphabet et L et M deux langages sur A.
- On peut définir les opérations suivantes :
  - L'**intersection** de L et M : $\mathbf{L \cap M}$ est formé par tous les mots sur A appartenant simultanément à L et à M.
  - La **réunion** de L et M : $\mathbf{L \cup M}$ est formé par tous les mots sur A appartenant à L ou à M (noté aussi L+M).
  - La **concaténation** de L et M : **LM** est formé par tous les mots sur A, obtenus en concanétant tous les mots de L avec tous les mots de M.
  - La **fermeture de Kleene** (ou **l'étoile**) de L : $\mathbf{L^*}$ est formé par tous les mots sur A qui sont des puissances de mots de L.

27

Opérations sur les langages

- La **fermeture de Kleene positive** (ou **l'étoile positive**) de L : $\mathbf{L^+}$ est formé par tous les mots sur A qui sont des puissances non nulles de mots de L.
- La **complémentation** de L : $\overline{L}$ est formé par tous les mots sur A n'appartenant pas à L.

- Formellement :
  - $L \cap M = \{s \in A^* \mid s \in L \text{ et } s \in M\}$.
  - $L + M = L \cup M = \{s \in A^* \mid s \in L \text{ ou } s \in M\}$.
  - $LM = \{s \in A^* \mid \exists u \in L, \exists v \in M : s = uv\}$.
  - $L^* = \{s \in A^* \mid \exists u_1, \ldots, u_n \in L : s = u_1 \ldots u_n\} \cup \{\varepsilon\}$
  - $L^+ = L^* \setminus \{\varepsilon\}$.

28

Opérations sur les langages

- $\overline{L} = \{s \in A^* \mid s \notin L\}$.

- On note L^n l'ensemble formé par les puissances n ^{ème} des mots appartenant à L (appelé **puissance n ^{ème}** de L).
- Quelques propriétés importantes :
  - L'intersection et la réunion sont commutatives, associatives et distributives, l'une par rapport à l'autre.
  - $L^0 = \{\varepsilon\}$ et $L^1 = L$, $L^2 = LL$, et ainsi de suite.
  - $L^* = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \ldots$
  - $L^+ = L^1 \cup L^2 \cup L^3 \cup \ldots$
  - $(LM)N = L(MN)$

29

Opérations sur les langages

- $L\{\varepsilon\} = \{\varepsilon\}L = L$.
- $L(M + N) = LM + LN$.
- $L(M \cap N) =$ (inclu)$LM \cap LN$.

- On peut aussi définir le **miroir d'un langage** L comme étant formé par les miroirs des mots appartenant à L. On obtient alors un autre langage, noté L^R. Formellement : $L^R = \{s^R \mid s \in L\}$.
- On a les propriétés suivantes :
  - $(L^R)^R = L$.
  - $(LM)^R = M^R L^R$.
  - $(L + M)^R = L^R + M^R$ et $(L \cap M)^R = L^R \cap M^R$.

30

Opérations sur les langages

- C^+ : langage des mots d'au moins un chiffre.

- Remarque :
  - Toutes les unités lexicales d'un langage de programmation peuvent être décrites en utilisant les trois opérations suivantes :
    - La concaténation.
    - La réunion.
    - La fermeture de Kleene.
- Exemple : $L(L + C)^*$ est l'ensemble de tous les identificateurs en Pascal.