Langages

1 Introduction
1-1-Définitions

- 1- Notation:
  - Symbole
  - Alphabet A: ensemble non vides de symboles
  - Mot : une suite de symboles
    - Longueur d'un mot |m|
    - ε: le mot vide de longueur 0.
    - A *: L'ensemble de tous les mots possibles sur un alphabet A.
- Définition:

Un langage L sur un alphabet A est un ensemble de mots (ou ensemble de chaînes) sur A. L est donc un sous-ensemble de A*, autrement dit $L \subseteq A^*$

donc L ∈ P(A*)

Exemples

- *A = {0, 1}* alors *L1 = {0, 00, 1, 01, 11, 10 } est un langage sur A*

Définir A et P(A*)*

- *An: ensembles des mots de longueur n.*
- Le langage neutre est celui dont le seul mot est la chaîne vide L = {ε}.
- Le langage vide est celui qui ne contient aucun mot, soit L= ∅.
- Un langage fini est un langage qui contient un nombre fini de mots.
- Un langage infini est un langage non vide et non fini.

1-2- Relations entre les mots

- Soit x et y sont deux mots sur A, alors:
  - x est un **préfixe** de y (*x ∈ Pref(y)*) si et seulement si $\exists z \in A^*$, $y = xz$;
  - x est un **suffixe** de y (*x ∈ Suff(y)*) si et seulement si $\exists$ $z \in A^*$, $y = zx$;
  - x est une **sous-chaîne** (ou **facteur**) $x \in Fact(y)$ si et seulement si $\exists$ w, z $\in A^*$, $y = wxz$.
  - Si $x \neq y$ (c'est-à-dire si $z \neq \varepsilon$), alors le préfixe ou le suffixe est dit **propre**.
- Remarque:
  - ε est préfixe, suffixe et sous-chaîne de toute chaîne. $Pref(x) \subseteq Fact(x)$ et $Suff(x) \subseteq Fact(x)$
- Exerices
- Soit *x=abbcc* un mot sur l'alphabet *A={a,b,c}* :
  - Donner l'ensemble *Pref(x)*
  - Donner l'ensemble *Suff(x)*

- Un **langage** L est dit posséder la propriété **préfixe** (resp. **suffixe**) si aucune chaîne de L n'est préfixe propre (resp. suffixe propre) d'une autre chaîne de L.
  - Par exemple, $L = \{a^i b \mid i \geq 0\} = \{b, ab, aab, aaab...\}$ est dit préfixe mais n'est pas suffixe. $L = \{a^n \mid n \in N\}$ n'est ni l'un ni l'autre.
- Remarque
  - le langage vide est absorbant pour la concaténation: $\emptyset L = \emptyset = \emptyset L$
  - Le langage neutre est élément neutre pour la concaténation des langages : $\{\varepsilon\}L = L = L\{\varepsilon\}$
  - La concaténation des langages est associative : $(L_1L_2)L_3 = L_1(L_2L_3)$

1-3 Opérations sur les langages

Soit A un alphabet. On définit sur les langages de $P(A^*)$ les opérateurs suivants : soient L et M deux langages sur A,

- Opérateurs ensemblistes classiques :
- **union** : $L \cup M = \{x \mid x \in L \text{ ou } x \in M\}$;
- **intersection** : $L \cap M = \{x \mid x \in L \text{ et } x \in M\}$;
- **différence** (ou exclusion) : $L \backslash M = L\text{-}M = \{x \mid x \in L \text{ et } x \notin M\}$;
- **complémentaire** sur A* : $Comp(L)= A^* \backslash L = \{x \mid x \in A^* \text{ et } x \notin L\}$;
- Opérateurs induits par la concaténation des mots :
- **produit** des langages : $LM = L \times M = \{xy \mid x \in L \text{ et } y \in M\}$;
  - Exemple: $L_1=\{a,bc\}$ et $L_2=\{de,f\}$ alors $L_1 \times L_2=\{ade, af, bcde, bcf\}$.
- **fermeture** de Kleene : $L^* = \cup_{i=0..\infty} L^i$ où $L^0 = \{\varepsilon\}$ et $L^n = LL^{n-1} = L^{n-1}L$;
- **fermeture positive** : $L^+= \cup_{i=1..\infty} L^i$.

2 Les langages rationnels

Soit A un alphabet. Les **langages rationnels sur A** (appelés aussi **langages réguliers**) sont les éléments de la classe Rat(A*) définie inductivement de la façon suivante :

1. $\emptyset$ est un langage rationnel $\in$ Rat(A*);
2. $\{\varepsilon\} \in$ Rat(A*);
3. $\forall a \in A$, $\{a\} \in$ Rat(A*);
4. Si $L_1 \in$ Rat(A*) et $L_2 \in$ Rat(A*) alors $L_1 \cup L_2 \in$ Rat(A*);
5. Si $L_1 \in$ Rat(A*) et $L_2 \in$ Rat(A*) alors $L_1 L_2 \in$ Rat(A*);
6. Si $L_1 \in$ Rat(A*) alors $L_1^* \in$ Rat(A*).

- Exemples (a,b,c)
- **Théorème des parties finies**: Toute partie finie L de A* est dans Rat(A*).

3- Les expressions rationnelles

Les **expressions rationnelles** sur A décrivent les **langages rationnels**. Elles sont définies de la façon suivante :

- $\emptyset$ est une expression rationnelle qui décrit le langage rationnel $\emptyset$;
- ε est une expression rationnelle qui décrit le langage rationnel {ε} ;
- pour tout a de A, a est une expression rationnelle qui décrit le langage rationnel {a} ;
- Si l_1 et l_2 sont des expressions rationnelles qui décrivent L_1 et $L_2 \in$ *Rat(A*)* alors :
  - $(l_1 \mid l_2)$ et $(l_1 + l_2)$ sont des expressions rationnelles identiques qui décrivent le langage rationnel $L_1 \cup L_2 \in Rat(A^*)$,
  - $(l_1 l_2)$ et $(l_1 . l_2)$ sont des expressions rationnelles identiques qui décrivent le langage rationnel $L_1 \times L_2 \in Rat(A^*)$,
  - $(l_1)^*$ est une expression rationnelle qui décrit le langage rationnel $(L_1)^*$.

Deux expressions rationnelles w_1 et w_2 sont dites équivalentes (ou par abus de langage "égales"), noté "$w_1 \equiv w_2$" (ou "$w_1 = w_2$"), si elles décrivent le même langage rationnel

- Propriétés
  - L'union $u|v = v|u$ (commutativité)
  - $u|u = u$ (idempotence)
  - Aiguilles