Ch 3 Expressions régulières

1. Mots & langages formels

- <u>Alphabet</u> : ensemble <u>fini</u> de symboles
- <u>Mot</u> (sur un alphabet) : suite <u>finie</u> de lettres de l'alphabet
- <u>Concaténation</u> : opération qui permet de former un mot en juxtaposant deux mots.
- <u>Facteur</u> : un facteur w d'un mot u est une suite de lettres adjascente dans u.

  facteur gauche : préfixe
  facteur droit : suffixe

$X = \{a, b, c, d\}$

$u = aaba$

$v = cbaaa$

$u.v = aabacbaaa$

$u.u = aabaaaba$

facteurs de u :

aa

aba

a

} facteurs de u

préfixe

suffixe

ab facteur de u (ni préfixe ni suffixe)

$babac$

$= \underline{ba}.\underline{bac}$

$= \underline{b}.aba.\underline{c}$

$= b.a.b.a.c$

} factorisat

$12 = 3 \times 4$

$= 2 \times 6$

$= 3 \times 2 \times 2$

$\vdots$

$X = \{ le, chat, est, mot, petit \}$ (token)

mot : le le chat le

le petit chat

chat chat chat chat

$X = \{$ mang-, -iez, -er-, ... $\}$

mot : mange er iez

er er er

alphabet
mot
concaténation

Langage (formel)

Def : Soit X un alphabet, on appelle X^* l'ensemble (infini) de tous les mots qu'on peut former sur X.
Un langage est un sous-ensemble de X^*.

"Un langage est un ensemble de mots"

Langage formel vs. langue naturelle

Français $\stackrel{?}{=}$ langage formel
(syntaxe)

alphabet : morphèmes
mots : phrases
langage : sous-ens. de toutes les combinai. sous de morphèmes qui sont des phrases bien formées.

"English as a Formal Language"

(CHOMSKY)

Description d'un langage (formel) ?

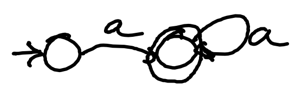
- liste des mots $\rightarrow$ possible seulement si le langage est fini.

$$X = \{a, b, c\}$$

$$X^* = \{a, aa, bc, abacbac, \dots\}$$

$$L_1 \subset X^* = \{a, aa, bbb\}$$

$$L_2 = \{a, aa, aaa, aaaa \dots\}$$

- expression régulière $L_2 = aa^*$
- automate 

- grammaire formelle $\begin{cases} S \rightarrow aS \\ S \rightarrow a \end{cases}$

moyens finis de représenter un ens. infini.

Humbolt

2. Opérations sur ~~les~~ langages.

2.1 Opérations ensemblistes.

Union
Intersection
Complément

$$X = \{a, b, c\}$$

$$L_1 = \{ab, ba\} \quad L_2 = \{a, aa, aaa\}$$

$$L_1 \cup L_2 = \{ab, ba, a, aa, aaa\}$$

$$L_1 \cap L_2 = \emptyset$$

2.2 Opérations avec la concaténation

"Produit de langages"

$X = \{a, b, c, d\}$

$L_1 = \{ab, bc, cd, da\}$ $\qquad L_2 = \{b, bb, bbb \dots\}$

$$L_1 \cdot L_2 = \{ab.b,\ ab.bb,\ ab.bbb \dots,$$
$$bc.b,\ bc.bb,\ bc.bbb \dots,$$
$$cd.b,\ cd.bb,\ \dots$$
$$da.b,\ da.bb \dots \}$$

$L_1 = \{le, la, les\}$ $\quad$ $L_2 = \{chien, chatte, oiseaux\}$

$L_1.L_2 = \{le.chien, le.chatte, le.oiseaux \ldots\}$

$L_1.L_2$ = ens. des mots formés en concaténant un mot de L_1 avec un mot de L_2.

mot-vide : ε $\quad$ mot qui ne contient aucune lettre.

$X = \{a, b\}$ $\quad$ $L_1 = \{\varepsilon, a, bbb\}$

$L_2 = \{bab, aba\}$

$L_1.L_2 = \{\varepsilon.bab, aba, abab, aaba, bbbbab, bbbaba\}$

(bab)

Propriétés de "."

Concaténation : non commutative

$u.v \neq v.u$

élément neutre :

$u.\varepsilon = \varepsilon.u = u$

mots = suite de lettres

Produit : non commutative

$L_1.L_2 \neq L_2.L_1$

élément neutre :

$L.X = L \qquad X = \{\varepsilon\}$

$L.\{\varepsilon\} = L$

élément absorbant :

$L.X = X \qquad X = \emptyset$

$L.\emptyset = \emptyset$

lges = ens de mots

$L_1 = \{a, ab\}$

$L_2 = \emptyset$

$L_1.L_2 = \emptyset$

ex. $L = \{a, bb, ccc\}$

$L.L = \{aa, abb, accc, bba, bbbb, bbccc, ccca, cccbb, cccccc\} = L^2$

$L^2.L = L^3 = \{aaa, aabb, aaccc, \dots\}$

$L = L^1$

$L.L = L^2$

$L.L.L = L^3$

$\vdots$

Convention : $L^0 = \{\varepsilon\}$

$x^2 = x.x$

$x^1 = x$

$x^0 = 1$ (clt neutre du produit)

2.3 Opérations rationnelles

Union Produit Etoile.

$$L^* = L^0 \cup L^1 \cup L^2 \cup \dots$$

$$\underset{def}{=} \{\varepsilon\} \cup L \cup L.L \cup L.L.L \ \dots\dots$$

"étoile de Kleene"