LIF15 – Théorie des langages formels

Sylvain Brandel

2015 – 2016

sylvain.brandel@univ-lyon1.fr

CM 2

ALPHABETS ET LANGAGES

Alphabets et langages
Alphabet

- Un <u>alphabet</u> est un ensemble <u>fini</u>, <u>non vide</u>, de symboles.

  On le note généralement Σ.

- Un <u>mot</u> sur un alphabet Σ est une suite finie d'éléments de Σ.

- On note Σ^* l'ensemble de <u>tous</u> les mots (y compris le mot vide) définis sur Σ.

Alphabets et langages
Alphabet

- <u>Longueur</u> : nombre de symboles d'un mot.

- Deux mots u et v sont égaux ssi
  - ils ont même longueur
  - $\forall$ i $\in$ {1, …, |u|} : $u_i = v_i$.

Alphabets et langages
Alphabet

- La <u>concaténation</u> de 2 mots u et v de Σ^* est un mot noté uv et défini par :
  - $u = u_1u_2\ldots u_n$, $v = v_1v_2\ldots v_n$ → $w = u_1\ldots u_nv_1\ldots v_n$
    - $\forall\, i \in \{1, |u|\}$ $(uv)_i = u_i$
    - $\forall\, i \in \{|u|+1, \ldots |u|+|v|\}$ $(uv)_i = v_{i-|u|}$

- <u>Propositions</u>
  - la concaténation est régulière à droite et à gauche
    - $wu = wv \Rightarrow u = v$
    - $uw = vw \Rightarrow u = v$
  - $|uv| = |u| + |v|$.

Alphabets et langages
Alphabet

- On appelle <u>facteur gauche</u> de w un mot u tel que uv = w.
- On appelle <u>facteur droit</u> un mot v tel que uv = w.
- On appelle <u>facteur</u> de w un mot u tel que il existe v et v' tels que vuv' = w.

- <u>Miroir</u> (Reverse) :
  - La fonction miroir $\_^R : \Sigma^* \rightarrow \Sigma^*$ est définie par récurrence :
    - w tq $|w| = 0$: $w^R = e^R = e$
    - w tq $|w| > 0$: $\exists\, a \in \Sigma$ tq $w = au$

      et $w^R = (au)^R = u^R a$

- <u>Propriété</u>
  - $\forall\, u, v \in \Sigma^* : (uv)^R = v^R u^R$

Alphabets et langages
Langage

- On appelle langage sur Σ tout ensemble de mots sur Σ

- Remarque de cardinalité
  - Σ fini
  - Σ^* infini dénombrable (rappel : dont on peut énumérer les éléments)
  - P (Σ^*) est infini non dénombrable

- Opérations sur les langages
  - $\cup$, $\cap$, $\neg$ (complément), \ $\Rightarrow$ comme d'habitude
    (complément : $\neg A = \Sigma^* \setminus A$)
  - Concaténation : $L_1 \subset \Sigma^*$, $L_2 \subset \Sigma^*$
    - $L = L_1.L_2$ ou L_1L_2
      est défini par $L = \{w \mid \exists\, w_1 \in L_1$ et $\exists\, w_2 \in L_2 : w = w_1w_2\}$
  - Clôture de Kleene (Kleene star) ou étoile de L
    - $L^* = \{w \in \Sigma^* \mid \exists\, k \in N, \exists\, w_1, w_2, \ldots, w_k \in L : w = w_1w_2\ldots w_k\}$

Représentation finie des langages

- Une description habituelle d'une certaine classe de langages est fournie par ce qu'on appelle les <u>expressions régulières</u> (ou <u>rationnelles</u>).
  - les éléments de base sont :
    - les singletons sur Σ
    - l'ensemble $\varnothing$
  - les opérations sont :
    - la concaténation de langages
    - la réunion de deux langages
    - la fermeture de Kleene

Représentation finie des langages

- Les expressions régulières / rationnelles constituent syntaxiquement un langage (que nous appellerons ER) de mots bien formés sur l'alphabet $\Sigma \cup \{(,), \emptyset, \cup, *\}$ tel que :

  ① $\emptyset$ et chaque lettre de Σ est une ER

  ② si α et β sont des ER, alors $(\alpha\beta)$ et $(\alpha \cup \beta)$ aussi

  ③ si α est une ER alors α^* aussi

  ④ ER est close pour ces propriétés

  (rien d'autre n'est une ER que les points ① à ③)

Représentation finie des langages

- <u>Exemple</u>
  - $\Sigma = \{a, b, c\}$
  - Ensemble des mots finissant par a :

$$L = (a \cup b \cup c)^* a$$

$$(= (a^* c \cup b)^* \, a)$$

- <u>Définition</u>
  - On appelle <u>Langage rationnel</u> (ou <u>régulier</u>) tout langage qui peut être décrit par une expression rationnelle.