Langages

Alphabet

Alphabet : Ensemble fini de symboles (lettres ou caractères), noté Σ

Nous utilisons des lettres minuscules pour les lettres de l'alphabet : $\Sigma_4 = \{a, b\}$

- Exemples :
- Σ_1=ASCII,
- Σ_2={0,1} (*alphabet binaire*),
- Σ_3={a,b,c, …,z}.
- Σ_4 = {if, then, else, a, b}

Mots

Mot (chaine) : Suite (séquence) finie et ordonnée, éventuellement vide, d'éléments de Σ

C'est une concaténation de lettres.

Exemples : "informatique", "langage", "THL", définis sur l'alphabet $\Sigma = \{a, b, c, \ldots, z\}$

Mots (chaines) sur l'αb $\Sigma_4 = \{a, b\}$

a

ab

bba

$abaa$

$aabbbaababa$

$u = ab$

$v = bbaaab$

$w = baba$

Mot vide

Le mot vide, noté ε, correspond à la suite de symboles vide.

Une chaine avec aucune lettre

Elément neutre de la concaténation :

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

$$\varepsilon abb = abb\varepsilon = abb$$

L'Opération * sur (Fermeture Transitive) Σ

Σ* : L'ensemble de tous les mots possibles formés à partir de l'alphabet Σ

Σ* est comptable infini (dénombrable)

Exemples :

$$\text{Pour } \Sigma_2 = \{0,1\} \text{ et } \Sigma_4 = \{a, b\}$$

$$\Sigma_2{}^* = \{\varepsilon, 0, 1, 00, 11, 01, 10, 110, 001, \ldots\}$$

$$\Sigma_4{}^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

L' Opération + sur Σ

Σ^+ : L'ensemble de tous les mots possibles à former à partir de l'alphabet Σ à l'exception de ε

Exemple :

$$\Sigma_4 = \{a, b\}$$

$$\Sigma_4^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

$$\Sigma_4^+ = \Sigma_4^* \setminus \{\varepsilon\}$$

$$\Sigma_4^+ = \{a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

Concaténation

- Juxtaposition de mots…

$$.: \Sigma^* \times \Sigma^* \rightarrow \Sigma^*$$

$$u, v \rightarrow .(u, v) = u.v \text{ noté seulement } uv$$

e.g. $u = a_1 a_2 \ldots a_n$ abaa

$v = b_1 b_2 \ldots b_m$ abb

$\text{uv} = a_1 a_2 \ldots a_n b_1 b_2 \ldots b_m$ abaa.abb = abaaabb

. est associative mais non commutative ;

i.e. en général, $uv \neq vu$ mais $(uv)w = u(vw)$

Σ^* est clos par rapport à la concaténation, i.e.

$$u, v \in \Sigma^* \text{ alors } uv \in \Sigma^*$$

Puissance $u^n = \underbrace{uu\cdots u}_{n}$

$$u^n = \begin{cases} u\,u^{n-1} & \text{si } n>0 \\ \varepsilon & \text{si } n=0 \end{cases}$$

Exemple :

$$u = ab; \quad u^0 = \varepsilon; \quad u^1 = u; \; u^2 = abab.$$

$$(ababa)^2 = ababaababa$$

$$(ababa)^0 = \varepsilon$$

Remarque : a^n est le mot composé de n occurrences de a (a^0 est le mot vide).

Inverse (Miroir) d'un Mot

Soit un Σ et $u \in \Sigma^*$ tel que $u = a_1 a_2 \ldots a_n$

$$\text{avec } \forall i \in \{1, \ldots, n\}, a_i \in \Sigma$$

L'inverse (image miroir) de u, noté $\widetilde{u}$, est défini par

$$\widetilde{u} = a_n a_{n-1} \ldots a_1$$

Définition récursive :

$$\widetilde{u} = \begin{cases} u \ \text{ si } u = \varepsilon \\ \widetilde{v} a \text{ si } u = av \text{ avec } a \in \Sigma \text{ et } \in \Sigma^* \end{cases}$$

Exemple : $u = aabbab$ alors $\widetilde{u} = babbaa$

Longueur d'un Mot

Le nombre de lettres qui composent un mot

$$| \ | : \Sigma^* \to \mathrm{N}$$

$$u \to |u|$$

Si $u = a_1 a_2 \cdots a_n$ alors $|u| = n$

Le mot vide est de longueur 0. $|\varepsilon| = 0$

Exemples :

$$|abba| = 4$$

$$|aa| = 2$$

$$|a| = 1$$

Le nombre d'occurrence d'une lettre

$$|u|_a : \Sigma^* \to \mathrm{N}$$

$$u \to |u|_a$$

est le nombre d'occurrences de *a* dans *u*

Exemples :

$$u = \text{information}$$

$$|u|_a = 1$$

$$|u|_0 = 2$$

$$|u|_i = 2$$

Longueur de la Concaténation

$$|uv| = |u| + |v|$$

Exemple :

$$u = ab, \quad |u| = 2$$

$$v = abbab, \quad |v| = 5$$

$$|uv| = |ababbab| = 7$$

$$|uv| = |u| + |v| = 2 + 5 = 8$$

Sous-Mots

Sous chaines :

Une sous-séquence de caractères consécutifs

Mots	Sous-mots
$\underline{aa}bab$	aa
$\underline{abba}b$	$abba$
$aab\underline{a}b$	a
$a\underline{baab}$	$baab$

Préfixe, Suffixe et Facteur

Soit un alphabet Σ et $u, v \in \Sigma^*$

$$-u \text{ est un } \textbf{préfixe} \text{ de } v \text{ ssi } \exists \mathrm{w} \in \Sigma^*$$

$$\text{tel que } v = uw$$

$$-u \text{ est un } \textbf{suffixe} \text{ de } v \text{ ssi } \exists \mathrm{w} \in \Sigma^*$$

$$\text{tel que } v = wu$$

$$-u \text{ est un } \textbf{facteur} \text{ de } v \text{ ssi } \exists \mathrm{w}_1, \mathrm{w}_2 \in \Sigma^*$$

$$\text{tel que } v = w_1 u w_2$$

Préfixe et Suffixe

$$w = ababa$$

Préfixes	Suffixes
ε	$ababa$
a	$baba$
ab	aba
aba	ba
$abab$	a
$ababa$	ε

$$w = uv$$

préfixe

suffixe

Préfixe et Suffixe Propre

u est un préfixe propre de v ssi u est un préfixe de v et $u \neq v$.

u est un suffixe propre de v ssi u est un suffixe de v et $u \neq v$.

Monoïde

Un ensemble E muni d'une opération interne, notée ., associative et possédant un élément neutre est un monoïde, noté $\mathrm{M} = (E,.)$

Exemples :

$$-\mathrm{M}_1 = (N,+)$$
$$-\mathrm{M}_2 = (N,\times)$$
$$-\mathrm{M}_3 = (\Sigma^*,.)$$

Langages

Un langage est un ensemble de mots formés sur l'αb Σ

Un langage est un sous-ensemble de Σ^* (fini ou infini) $\left(L \subseteq \Sigma^*\right)$

Exemple de langages de : Si

$$\Sigma = \{a, b\} \text{ alors } \Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \ldots\}$$

et les ensembles

$$L_1 = \{\varepsilon\}$$

$$L_2 = \{a, aa, ab, ba, bb\}$$

$$L_3 = \{\varepsilon, ab, aaa, ababa, bbaba, aaaaaa\}$$

Sont des langages finis

Langages

Un langage L sur Σ est une partie de Σ^*

Exemple : $\Sigma_4 = \{a, b\}$

L1={aa, abba, bba }est un langage fini

L2 = $\left\{ u \in \Sigma^* \middle| \; u = au'b \; et \; u' \in \Sigma^* \right\}$

= {ab, aaaab, a…….b} est un langage infini

Pb : Comment décrire un langage d'une manière formelle pour faciliter son traitement par un ordinateur ?

Langages -Remarques-

Deux langages différents $\varnothing = \{\ \} \neq \{\varepsilon\}$

$\{\ \}$: l'ensemble /langage vide, ne contenant aucune chaine

$\{\varepsilon\}$: Un langage contenant une seule chaine, la chaine vide.

Taille de l'ensemble $\left|\{\ \}\right| = \left|\varnothing\right| = 0$

Taille de l'ensemble $\left|\{\varepsilon\}\right| = 1$

Longueur du mot $\left|\varepsilon\right| = 0$

Théorie Standard des ensembles

Conditionnel : $A = \{ x \mid x \in N , f(x)=0 \}$

Union : $A \cup B$

Intersection : $A \cap B$

Complément : $\overline{A}$

Produit Cartésien : $A \times B$

Parties d'ensemble : $\mathcal{P}(A)$

On utilise cette théorie pour représenter les langages

Parties d'ensembles

Formellement: $\mathcal{P}(A) = \{ S \mid S \subseteq A\}$

Exemple: $A = \{x,y\}$

$\mathcal{P}(A) = \{ \{\} , \{x\} , \{y\} , \{x,y\} \}$

Noter la différence des tailles :

$|\mathcal{P}(A)| = 2^{|A|}$

$|A \times A| = |A|^2$

Un exemple de Langage (Conditionnel)

Un langage infini $$L = \{a^n b^n \mid \; n \geq 0\}$$

Les mots suivants :

$$\left.\begin{matrix} \varepsilon \\ ab \\ aabb \\ aaabbb \\ aaaabbbb \end{matrix}\right\} \in L$$

et $aab, abab \notin L$

Autres exemples de langages (Conditionnel)

$$-L_1 = \left\{ x \in \{0,1\}^* \middle| x \text{ contient un nombre pair de 0'}s \right\}$$

$$-L_2 = \left\{ x \in \{0,1,...,9,.\}^* \middle| x \text{ est un nombre réel à longueur finie} \right\}$$

$$-L_3 = \left\{ x \in \{a,b,..., z\}^* \middle| \; x \, e\text{st mot réservé Pascal} \right\}$$

$$= \{if\, , begin, while, ...\}$$

$$-L_4 = L_3 \bigcup \{\, (,\,),\, .,\, :,\, ,\, ,\, \ldots \} \text{U} \, \{\text{identificateurs Légaux Pascal}\}$$

$$-L_5 = \left\{ x \in L_4{}^* \middle| \begin{array}{l} \text{x est un programme Pascal} \\ \text{syntaxiquement correct} \end{array} \right\}$$

Opérations sur les Langages

L'ensemble des opérations usuelles $L_1, L_2 \in \Sigma^*$

- Union $L_1 \bigcup L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ ou } u \in L_2\}$
  - Commutative
  - Associative
  - Possède élément neutre : ensemble vide ϕ

- Intersection $L_1 \bigcap L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ et } u \in L_2\}$
  - Commutative
  - Associative
  - Possède élément neutre $\sum^*$

- Différence $L_1 \setminus L_2 = \{u \in \Sigma^* \mid u \in L_1 \text{ et } u \notin L_2\}$

- Complémentaire $\overline{L} = \Sigma^* \setminus L$

Opérations sur les Langages

$$\{a, ab, aaa\} \cup \{ba, ab\} = \{a, ab, ba, aaa\}$$

$$\{a, ab, aaa\} \cap \{ba, ab\} = \{ab\}$$

$$\{a, ab, aaa\} \setminus \{ba, ab\} = \{a, aaa\}$$

$$\overline{\{b, ab\}} = \{\varepsilon, a, aa, ba, bb, aaa, \ldots\}$$

Inverse (Miroir)

Définition : $L^R = \{w^R \mid w \in L\}$

Exemples : $\{ab, aba, abab\}^R = \{ba, aba, baba\}$

$$L = \{a^n b^n \mid n \geq 0\}$$

$$L^R = \{b^n a^n \mid n \geq 0\}$$

Concaténation

Définition : $L_1 L_2 = \{uv \mid u \in L_1 et v \in L_2\}$

$$\{a, ab, ba\}\{b, bb\}$$

Exemple :

$$= \{ab, abb, abb, abbb, bab, babb\}$$

Attention

Ne pas confondre la concaténation des langages avec le produit cartésien d'ensembles.

Concaténation: $L_1.L_2 = \{uv | u \in L_1 \text{ et } v \in L_2\}$

Produit : $L_1 \times L_2 = \{(u,v) | u \in L_1 \text{ et } v \in L_2\}$

Par exemple, si $L_1 = \{0,00\}$

$$- L_1.L_1 = \{0,000,0000\} \text{avec } |L_1.L_1| = 3$$

$$- L_1 \times L_1 = \{(0,0),(0,00),(00,0),(00,00)\}$$

$$\text{avec } |L_1 \times L_1| = 4$$

Itération

Définition:

$$L^n = \underbrace{LL \cdots L}_{n}$$

$$\{a, bb\}^3 = \{a, bb\}\{a, bb\}\{a, bb\} =$$
$$\{aaa, aabb, abba, abbbb, bbaa, bbabb, bbbba, bbbbbb\}$$

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

Cas spécial :

$$\{a, bba, aaa\}^0 = \{\varepsilon\}$$

D'autres Exemples

$$L = \{a^n b^n : n \geq 0\}$$

$$L^2 = \{a^m b^m a^n b^n : m, n \geq 0\}$$

$$abaaabbb \in L^2$$

Fermeture (Kleene *)

$$L^* = \bigcup_{i \geq 0} L_i = L^0 \cup L^1 \cup L^2 \cdots$$

l'opération étoile (ou fermeture par étoile, ou fermeture de Kleene, ou fermeture itérative) du langage L

Exemple :

$$\{a, ab\}^* = \left\{ \begin{array}{l} \varepsilon, \\ a, ab, \\ aa, aab, aba, abab, \\ aaa, aaab, aaba, aabab, \ldots \end{array} \right\}$$

Fermeture Positive

$$L^{+} = \bigcup_{i\geq 1} L_i = L^1 \cup L^2 \cup \cdots$$

$$= L^{*} \setminus \{\varepsilon\} \quad \textbf{et} \;\; L^{+} = LL^{*} = L^{*}L$$

La fermeture positive du langage A.

Exemple :

$$\{a, ab\}^{+} = \left\{ \begin{array}{l} a, ab, \\ aa, aab, aba, abab, \\ aaa, aaab, aaba, aabab, \ldots \end{array} \right\}$$