Soit V un alphabet

Ensemble des mots sur $V = V^*$; mot vide $= \varepsilon$;

Soit $w \in V^*$ longueur $w = |w|$; $|\varepsilon| = 0$

$u \in V^*$ $\quad u[i] = a_i$ avec $u = a_1 a_2 \dots a_n$

$a_1, a_2, \dots, a_n \in V$;

$u[i..j]$ le facteur de u compris entre le $i^{ème}$ et $j^{ème}$ caractère de u.

$|w|_a$: nombre occurence de a dans w

$|w|_a = card\{i \in \mathbb{N} \text{ tq } w[i] = a\}$.

Concatenation de mot soit $u, v \in V^*$ $\begin{cases} u.v = a_1 \dots a_n b_1 \dots b_m \\ uv \end{cases}$

la concatenation est associative

$\forall u, v, w \in V^*$

$$(uv)w = u(vw)$$

ε est un element neutre :

$\forall u \in V^*$

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

Par reccurence :

$$w^0 = \varepsilon \text{ et } \forall n \in \mathbb{N}, \ w^{n+1} = w w^n$$

Def:

Facteur, un mot est un facteur d'un mot $w \in V^*_1$
si $\exists \alpha, \beta \in V^*$ tq $w = \alpha u \beta$
si $\alpha = \varepsilon$:

$w = u\beta$; u est un préfixe (ou facteur gauche) de w

si $\beta = \varepsilon$:

$w = \alpha u$; u est un suffixe (ou facteur droite) de w

Lemme : $u, v, x, y \in V^*$ et $a \in V$

1. $|uv| = |u| + |v|$

2. $|uv|_a = |u|_a + |v|_a$

3. $\forall i \leq |uv|$: $(uv)[i] = u[i]$ si $i \leq |u|$
   $v[i - [u]]$ sinon

4. $(xu = yv$ et $|x| < |y|) \Rightarrow x$ est préfixe de y et v est suffixe de u

5. $(xu = yv$ et $|x| = |y|) \Rightarrow x = y$ et $u = v$

6. $|u| = 0$ ssi $u = \varepsilon$

7. $|u| \geq n$ ssi $\exists x, y \in V^*$ tq $u = xy$ et $|x| = n$ et $|y| \geq 0$

Lemme 5 Si un sous ensemble X de V^* satisfait :

$$\boxed{\begin{array}{l} (i)\ \varepsilon \in X \\ (ii)\ \forall u \in X, \forall a \in V : au \in X \\ \text{Alors } X = V^* \end{array}}$$

Preuve : But $\forall u \in V^* \Rightarrow u \in X$

Recurrence sur la longueur :

Soit $u \in V^*$ et $|u| = 0$ $\Rightarrow u = \varepsilon$ donc $u \in X$ (i)

Supposons :

$|u| = n \Rightarrow u \in X$

Soit $v = ax$ avec $a \in V$ et $x \in X$ (donc $x \in V^*$)
(car $|x| = n \Rightarrow x \in X$ hypo)

$|v| = n+1 \Leftrightarrow |ax| = n+1$ or

(ii) $\Rightarrow ax \in X$ si $a \in V$ et $x \in X$

Donc initialisé et vraie au rang $n+1$ ($P(n) \Rightarrow P(n+1)$)

Corollaire 6 (Preuves inductives) :

Recurrence

Soit $\mathcal{P}$ une propriété définie sur V^* et telle que :

- $\mathcal{P}(\varepsilon)$ et (rang 0)
- si u satisfait $\mathcal{P}$, alors il en va de même pour au $\forall a \in V$
(rang n) (rang n+1)

Alors $\mathcal{P}$ est vraie $\forall u \in V^*$

Preuve : Soit $X = \{ u \in V^* \text{ tq } \mathcal{P}(u) \}$ vérifiant (i) et (ii) du L.5

Par conséquent $X = V^*$, donc $\forall u \in V^*$ satisfait $\mathcal{P}(u)$.

miroir, soit $w \in V^*$ son miroir est noté w^R :

$$w = w_1 \dots w_n \quad ; \quad w^R = w_n \dots w_1$$

Déf $\boxed{w \text{ est un palindrome ssi } w = w^R}$

Def : Un langage formel sur un alphabet V est une partie de V^*. (ensemble de mot de V^*)

- Pour tout alphabet V, les ensembles V^*, $\{\varepsilon\}$ et $\emptyset$ sont des langages sur V. langage vide

Opération sur les langages : (opération sur les ensembles classique :) "$\cup$", "$\cap$", "C", "/" et la concatenation des langages (définie par extension de la concatenation des mots)

Concatenation langage : L et L' deux langages sur un alphabet V,

$$LL' = \{uv \text{ tq } u \in L \text{ et } v \in L'\}$$

l'itération de cette opération sur le langage L permet d'obtenir

$$L^0, L^1, L^2, \dots$$

$L^0 = \{\varepsilon\}$ et $\forall n \in \mathbb{N}$, $L^{n+1} = LL^n$

Fermeture de Kleene (ou étoile) du langage L, le langage :

$$\boxed{L^* = \bigcup_{n \geq 0} L^n}$$

$\Leftrightarrow \{u \,/\, \exists k \geq 0 \text{ et } u_1 \dots u_k \in L \text{ tq } u = u_1 \dots u_k\}$

Theorème de Kleene :

L'ensemble rationnels sur un alphabet A est par definition le plus petit ensemble de parties de A^* contenant les singletons (parties reduites à un seul élément) et l'ensemble vide, et fermé par les opérations d'union, produit et étoile. Cet ensemble est noté $\mathrm{Rat}\ A^*$.

On appelle langage reconnaissable sur un alphabet A tout langage qui peut être reconnu par un automate fini sur A. L'ensemble des langage reconnaissables est noté $\mathrm{Rec}\ A^*$.

Theorème de Kleene - Sur un alphabet fini A, il y a égalité entre langages rationnels et langages reconnaissables. En d'autres termes, on a :

$$\mathrm{Rat}\ A^* = \mathrm{Rec}\ A^*$$

Ainsi L^* est constitué des mots de V^* qui s'obtiennent par la concaténation d'un nombre quelconque de mots de L.

étoile stricte :

$$\boxed{L^+ = \bigcup_{n \geq 1} L^n}$$

Note : Quelque soit le langage L, sa fermeture de Kleene L^* contient ε ($\{\varepsilon\} = L^0 \subseteq L^*$)

$$\varepsilon \in \emptyset^*$$

Def (clôture par la concaténation)

Un langage $L \subseteq V^*$ est clos par la concaténation si :

$$\forall x, y \in V^* : x, y \in L \Rightarrow xy \in L$$

Lemme 10 : Pour tous langage L :

L^* est le plus petit langage contenant L et clos par la concaténation

PREUVE :

1) L^* clos par la concaténation :

si $x, y \in L^*$ alors $\exists n, p \in \mathbb{N}$ tel que $x \in L^n$ et $y \in L^p$

Autrement dit :

$\exists x_1, \dots, x_n \in L$ et $y_1, \dots, y_p \in L$ tel que :

$$x = x_1 \dots x_n \text{ et } y = y_1 \dots y_p$$

Donc $xy = x_1 \dots x_n y_1 \dots y_p \in L^{n+p} \subseteq L^*$

$L^n \cdot L^p = L^{n+p}$

2) Si L' est un langage contenant L ($L \subset L'$) et clos par la concaténation on a :

$\forall n \in \mathbb{N} : \forall x_1, \dots x_n \in L, \ x_1, \dots x_n \in L'$

D'où : $L^n \subseteq L'$ $\forall n \in \mathbb{N}$, et donc $L^* = \bigcup_n L^n \subseteq L'$

Lemme 11 : Soit A, B et C trois langages sur l'alphabet V

Alors :

- $(AB)C = A(BC)$
- $(A \cup B)C = AC \cup BC$
- $(A \cap B)C \subset AC \cap BC$ (inclusion réciproque souvent fausse)
- $A \subseteq B \Rightarrow A^* \subseteq B^*$
- $(A^*)^* = A^*$
- $(A \cup B)^* = (A^* B^*)^*$

Le miroir d'un langage L, noté L^R est l'ensemble des miroirs des mots de L.

Autrement dit : $L^R = \{ w^R \text{ tq } w \in L \}$.

Langage regulier

Soit $\mathcal{P}(V^*)$ l'ensemble des parties de V^*, càd l'ensemble des langages d'alphabet V.

L'union, la concaténation des langages et la fermeture de Kleene sont des opérations d'arités respectives 2, 2 et 1 sur $\mathcal{P}(V^*)$. La classe des langages réguliers est définie inductivement à partir de ces opérations et des atomes $\emptyset$, $\{\varepsilon\}$, $\{a\}$, $a \in V$.

Def : (Langage regulier) L'ensemble REG des langages reguliers sur un alphabet V est le plus petit des sous-ensembles X de $\mathcal{P}(V^*)$ qui satisfont :

I

1. $\emptyset \in X$, $\{\varepsilon\} \in X$ et $\{a\} \in X$ pour chaque $a \in V$;
2. si $A, B \in X$ alors $A \cup B$, AB et A^* sont dans X

$\Updownarrow$

II

Autrement dit, REG est le plus petit ensemble de langages sur V qui contienne le langage vide (1) et tous les langages reduit à un mot de longueur ≤ 1 ($\Leftrightarrow$ 0 ou 1 : $\{\varepsilon\}$ et $\forall a \in V$), et qui soit clos pour l'union la concatenation et fermeture de Kleene (2.)

$\Updownarrow$

III

Les langages reguliers sont les langages obtenus à partir des «atomes» $\emptyset$, $\{\varepsilon\}$ et $\{a\}$, $a \in V$ par un nombre fini d'applications des trois opérations précitées.

Expression regulières:

Def (Expression Regulière): L'ensemble ER des expressions regulières sur un alphabet V est le plus petit des langages L sur l'alphabet $V \cup \{+, *, \emptyset, \varepsilon\}$ qui satisfont:

1. $\emptyset \in L$, $\varepsilon \in L$ et $a \in L$ pour chaque $a \in V$;
2. si $\alpha, \beta \in L$, alors $\alpha + \beta$, $\alpha\beta$ et α^* sont dans L;

A chaque expression regulière ξ sur V on associe un langage sur V selon les modalités suivantes:

1. $L(\emptyset) = \emptyset$; $L(\varepsilon) = \{\varepsilon\}$;
2. $L(a) = \{a\}$ pour tout $a \in V$;
3. $L(\alpha + \beta) = L(\alpha) \cup L(\beta)$;
4. $L(\alpha\beta) = L(\alpha) L(\beta)$;
5. $L(\alpha^*) = L(\alpha)^*$;

Le langage ainsi associé à une expression regulière ξ est dit dénoté par ξ.

Exemples:

- $\{a^n b^p, n, p \geq 0\}$ est dénoté par $a^* b^*$.
- $\{aabc^n, n \geq 0\}$ est dénoté par $aabc^*$
- $\{a_1, \dots, a_k\}$ est dénoté par $(a_1 + \dots + a_k)^*$;
- $(a+b)^* a (a+b)^*$ dénote le langage $\{w \in \{a,b\}^*$ tq $|w|_a \geq 1\}$.

Théorème : Un langage est régulier si, et seulement si, il est dénoté par une expression régulière. Autrement dit

$$REG = \{ L(\alpha) \mid \alpha \in ER \}$$

Preuve. $\subseteq$: il s'agit