

Théorie des langages

lien epita.lrde.epita.fr

chercher THL.

chercher pdf PTAPG, Parsing Techniques : A Practical Guide

Langage artificiel et naturel. Langage artificiel bcp plus simple que la langue naturelle.
Langage est un ensemble de mots. Un mot est une suite de symboles. Finalement, sinon on peut pas les traiter. Par exemple, « WHILE », est un symbole.

$\Sigma_1 = a, b$
u mots sur $\Sigma_1 : u = u_0 u_1 u_2 \dots u_n$
 $u_0, u_1, u_2, \dots, u_n \in \Sigma$

L est un lg $\Leftrightarrow$ L est un ensemble de mots.

Σ^* est l'ensemble des mots sur Σ .

$u \in \Sigma^*$

$L \subset \Sigma^*$

$L = \{u \in \Sigma^* \mid u \text{ code un nombre premier}\}$, L est le langage qui code un nombre premier

Soit ϵ le mot vide.

$(\Sigma^*, \cdot)$ est une LCI (stable), ϵ mots neutre $\epsilon u = u \epsilon = u$, est associatif : $u \cdot (v \cdot w) = (u \cdot v) \cdot w = u \cdot v \cdot w$
C'est un monoïde libre.

$u^n = \overbrace{u \dots u}^{n \text{ fois}}, u \in \Sigma^* \text{ et } n \in \mathbb{N}$
 $u^0 = \epsilon$

$u \ll_p v \Leftrightarrow v = u \cdot w, w \in \Sigma^* \text{ (p : prefixe)}$

$u \ll_f v \Leftrightarrow v = w \cdot u \cdot w', w \in \Sigma^* \text{ (f : facteur)}$

$u \ll_{\text{sous mots}} v \Leftrightarrow u \text{ est sous suite de } v$.

Ces relations ne sont pas totales.

relation d'ordre sur l'alphabet :

$a \ll_{rad} v \Leftrightarrow |u| < |v|$
 $\Leftrightarrow |u| = |v| \ \& \ u \ll_{alp} v$
 $a \ll b \ll aa \ll ab \ll ba$

$a \ll_{alp} v \Leftrightarrow u_0 < v_0$
 $\Leftrightarrow u_0 = v_0 \ \& \ u' \ll_{alp} v'$ avec $u = u_0 u', v = v_0 v'$

$d(u, v) = |u| + |v| - 2|pgpc(u, v)|$ (plus grand prefixe commun)

Recherche du mot le plus proche, on définit d'abord les opérations valides.

Exemple, opération : insérer, supprimer, échanger. La distance entre chien et chat est de 3 : s n, e e t, e i a.

concaténation des langages :

$L_1 \cdot L_2 = \{u \cdot v \mid u \in L_1, v \in L_2\}$

$\emptyset \cdot L = \{u \cdot v \mid u \in \emptyset, v \in L\} = \emptyset$. $\emptyset$ est absorbant

$\{\epsilon\} \cdot L = \{u \cdot v \mid u = \epsilon = L, v \in L\} = L$. $\{\epsilon\}$ est neutre

$L^n = \overbrace{L \dots L}^{n \text{ fois}}, u \in L$

$$L^0 = \epsilon$$

$$L = a, b \text{ et } L' = c, d$$

$$L^2 = \{aa, ab, ba, bb\}$$

$$L.L' = \{ac, ad, bc, bd\}$$

$$L \cup L' = L + L' = L|L'$$

$$L \cap L'$$

$$L \setminus L'$$

$$\bar{L}, \text{ partie complémentaire de } L$$

3 induction de base :

$$- a, b, \dots$$

$$- \emptyset$$

$$- \epsilon$$

operation noble :

$$- .$$

$$- \cup$$

$$- * (\text{etoile de Kleene}). (L^* = L^0 \cup L^1 \cup \dots \cup L^n \cup \dots = \cup_{n \in \mathbb{N}} L^n)$$

nombre $N = D^*.D$ avec $D = \{0\} \cup \{1\} \dots \cup \{9\}$

nombre signé : $S = (\{\epsilon\} \cup \{+\} \cup \{-\}).D^*.D$

nombre flottant : $F = S(\{\epsilon\} \cup (\{.\}.D^*)).(\{\epsilon\} \cup ((\{e\}) \cup \{E\}).S))$

expression rationnel

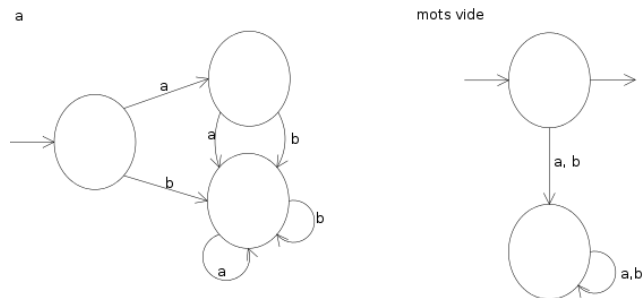
$$\alpha^+ = \alpha^* \alpha$$

$$\alpha^? = \epsilon + \alpha$$

$$\alpha^{\leq n} = \epsilon + \alpha + \alpha^2 + \dots + \alpha^n$$

$$\alpha^{\geq n} = \alpha^n + \alpha^{n+1} + \dots$$

$$\alpha^{n \dots m} = \alpha^n + \alpha^{n+1} + \dots + \alpha^m$$



$$A = (Q, q_0, Q_f, \Sigma_1, \delta)$$

- Q, q, q', q_0 ensemble d'états fini
- $q_0 \in Q$ état initial
- $Q_f \subset Q$ états finaux
- Σ_1 un alphabet
- $\delta : q, a \mapsto q'$
- $Q^* \Sigma_1 \rightarrow Q$

A est un automate deterministe. (une seul transition d'une lettre au max)

$A = (Q, Q_i, Q_f, \Sigma_1, \delta)$ est nondéterministe.

$$\delta \subset Q \cdot \Sigma_1 \cdot Q$$

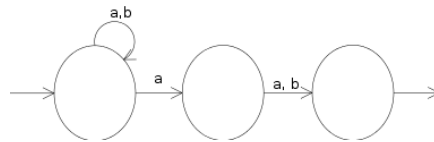
(présentation alternative :

$$\delta : Q \cdot \Sigma_1 \rightarrow 2^Q)$$

(non deterministe est plus général que déterministe)

$$\delta \subset Q \cdot (\Sigma_1 \cup \{\epsilon\}) \cdot Q$$

automate $(a|b)^*a(a|b)$



Lemme de pompage

$$q \vdash^a q'$$

$$\frac{A \quad B}{C} \text{ si A et B alors je pose C}$$

$$\frac{q \vdash^u q' \quad q' \vdash^v q''}{q \vdash^{uv} q''} \quad u \in L(A) \Leftrightarrow q_0 \vdash^u q_f$$

L, L' rationnels

- $L \cup L', L.L', L^*$
- Préfixe L
- Suffixe L
- Facteur L
- Sous-mots L
- $\bar{L}$
- $L \cap L'$

le resultat de ces opérations reste rationnel

$P \rightarrow S \vee C$

$S \rightarrow \text{il}$

$S \rightarrow \text{elle}$

$V \rightarrow \text{manger} \mid \text{boit}$

$C \rightarrow D \ N$

$D \rightarrow \text{des} \mid \text{les}$

$N \rightarrow \text{pommes} \mid \text{poires}$

Pour définir une grammaire, on a besoin de vocabulaire, symboles terminaux et non terminaux, de règles, couple (A, α) , $A \in N$ et $\alpha \in V^*$

$$\left. \begin{array}{l} T = \{\text{il}, \text{elle}, \text{mange}, \dots\} (= \Sigma) \\ N = \{P, S, V, C, \dots\} \end{array} \right\} V = N \cup T, T \cap N = \emptyset$$

$$G = (T, N, S, R), S \in N$$

T , ensemble des symboles terminaux

N , ensemble des symboles non terminaux

S , point d'entrée

R , ensemble des règles

G , la grammaire

R, T et N doivent être finis.

Une grammaire (syntagmatique) c'est :

$$G = (T, N, S, R)$$

T , ensemble fini de symboles terminaux

$a, b, c \dots$

N , ensemble fini des non terminaux

$A, B, C \dots$

$$V = T \cup N$$

$$T \cap N = \emptyset$$

proto mots / proto phrases

$$\alpha, \beta \dots \in V^* \text{ (et parfois } \epsilon \text{)}$$

R, ensemble fini de règles

(S, ϵ) ou (α, β) $\alpha, \beta \in V^*$

exemple :

$A \rightarrow aABC \mid abC$

$CB \rightarrow BC$

$bB \rightarrow bb$

$bC \rightarrow bc$

$cC \rightarrow cc$

$\underline{A} \vdash^2 \underline{abC} \vdash^5 \underline{abc}$

$A \vdash^1 \underline{aABC} \vdash^2 \underline{aabCBC} \vdash^3 \underline{aabBCC} \vdash^4 \underline{aabbCC} \vdash^5 \underline{aabbCC} \vdash^6 \underline{aabbcc}$

derivation :

$$\frac{\alpha \rightarrow \beta \in R}{\gamma\alpha\delta \vdash \gamma\beta\delta} \quad \gamma\delta \in V^* \frac{}{\alpha \vdash^* \alpha} \quad \frac{\alpha \vdash^* \beta \quad \beta \vdash^* \alpha}{\alpha \vdash^* \gamma} \quad \frac{\alpha \vdash \beta}{\alpha \vdash^* \beta}$$

$$\frac{S \rightarrow aS \in R}{S \vdash^* aS} \quad \frac{S \rightarrow aS \in R}{aS \vdash^* aaS} \quad \frac{S \rightarrow b \in R}{S \vdash^* b}$$

$$\frac{S \vdash^* aS \quad aS \vdash^* aaS}{S \vdash^* aaS} \quad \frac{S \vdash^* b}{S \vdash^* b}$$

$$\frac{S \vdash^* aaS \quad S \vdash^* b}{S \vdash^* aab}$$

$$L(G) = \{u \in T^* \mid S \vdash^* u\}$$

Sentence $\rightarrow$ Name | List End

Name $\rightarrow$ ceriel | dick | noam

List $\rightarrow$ Name — Name ',' List

',' Name End $\rightarrow$ and Name

$\Rightarrow$ 8 règles, 5 terminaux.

$\Rightarrow$ La grammaire crée une liste de prenom séparer de virgule.

$S \rightarrow N \mid L$

$N \rightarrow c \mid d \mid n$

$L \rightarrow E \mid NCL$

$CE \rightarrow \& E$

$\& E \rightarrow \& N$

$C \rightarrow ','$

$\Rightarrow$ grammaires Générales

$$\alpha \rightarrow \beta \quad \alpha, \beta \in V^+$$

$\Rightarrow$ grammaires sensibles au contexte

$$\gamma\alpha\delta \rightarrow \gamma\beta\delta \quad \gamma, \delta \in V^*$$

$\Rightarrow$ grammaires monotones

$$\alpha \rightarrow \beta \quad \alpha, \beta \in V^+ \mid |\alpha| \leq |\beta|$$

$S \rightarrow c \mid d \mid n \mid N$

$L \rightarrow cL' \mid dL' \mid nL'$

$L' \rightarrow ', ' \mid \&c \mid \&d \mid \&n$

$\Rightarrow$ grammaire relationnelle

$$\left. \begin{array}{l} A \rightarrow Bu \\ \rightarrow u \end{array} \right\} \text{lineaire a gauche}$$

$$\left. \begin{array}{l} A \rightarrow uB \\ \rightarrow u \end{array} \right\} \text{ linéaire à droite}$$