LifLF– Théorie des langages formels

Sylvain Brandel

2016 – 2017

sylvain.brandel@univ-lyon1.fr

CM 10

ALGÉBRICITÉ

Propriétés des langages algébriques

- Pour montrer qu'un langage est algébrique, on peut :
  - soit définir une grammaire algébrique qui engendre ce langage,
  - soit définir un automate à pile qui l'accepte.

- Il est également possible d'utiliser les propriétés de stabilité de la classe des langages algébriques

Propriétés des langages algébriques
Propriétés de stabilité

- Théorème

  La classe des langages algébriques est stable par les opérations d'union, de concaténation et d'étoile de Kleene.

- Preuve

  Soient deux grammaires $G_1 = (V_1, \Sigma_1, R_1, S_1)$ et $G_2 = (V_2, \Sigma_2, R_2, S_2)$, avec $V_1 \cap V_2 = \varnothing$. (On renomme éventuellement les non-terminaux.)

  La preuve (constructive) consiste à :

  - construire une grammaire G à partir de G_1 et G_2 validant les propriétés de stabilité,
  - montrer que $L(G) = L(G_1)$ op $L(G_2)$ (op $\in \{\cup, .\}$) et $L(G) = L(G_1)^*$.

Propriétés des langages algébriques
Propriétés de stabilité

- Preuve

(a) Union

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup V_2 \cup \{S\}$ où $S \notin V_1 \cup V_2$ (renommage éventuel)
- $\Sigma = \Sigma_1 \cup \Sigma_2$
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1 \mid S_2\}$

(b) Concaténation

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup V_2 \cup \{S\}$ où $S \notin V_1 \cup V_2$ (renommage éventuel)
- $\Sigma = \Sigma_1 \cup \Sigma_2$
- $R = R_1 \cup R_2 \cup \{S \rightarrow S_1S_2\}$

(c) Opération étoile

Soit $G = (V, \Sigma, R, S)$ avec :

- $V = V_1 \cup \{S\}$ où $S \notin V_1$ (renommage éventuel)
- $\Sigma = \Sigma_1$
- $R = R_1 \cup \{S \rightarrow S_1S \mid e\}$

Propriétés des langages algébriques
Propriétés de stabilité

- <u>Remarque</u>

  Contrairement à la classe des langages rationnels, la classe des langages algébriques n'est pas stable par intersection et complémentation.

- <u>Théorème</u>

  L'intersection d'un langage rationnel et d'un langage algébrique est algébrique

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Théorème (lemme de la double étoile)

  Soit L un langage algébrique.

  Il existe un nombre k, dépendant de L, tel que tout mot $z \in L$, $|z| \geq k$, peut être décomposé en $z = uvwxy$ avec :

  (*i*) $|vwx| \leq k$

  (*ii*) $|v| + |x| > 0$ (ie. $v \neq e$ ou $x \neq e$)

  (*iii*) $uv^nwx^ny \in L, \forall\, n \geq 0$ (d'où l'appellation de double étoile : v^n et $x^n = v^*$ et x^*)

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Pour le montrer on utilise la forme normale de Chomsky.

- <u>Définition</u>

  Une grammaire algébrique G = (V, $\sum$, R, S) est sous forme normale de Chomsky si chaque règle est de la forme :

  $A \rightarrow BC$ avec B, C $\in V - \{S\}$

  ou $A \rightarrow \sigma$ avec $\sigma \in \sum$

  ou $A \rightarrow e$

- <u>Théorème</u>

  Pour toute grammaire algébrique, il existe une grammaire sous forme normale de Chomsky équivalente.

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- <u>Lemme</u>

  Soit G = (V, ∑, R, S) une grammaire algébrique sous forme normale de Chomsky.

  Soit $S \Rightarrow_G^* w$ une dérivation de $w \in \Sigma^*$ dont l'arbre de dérivation est noté T.

  Si la hauteur de T est n alors $|w| \leq 2^{n-1}$.

- <u>Corollaire</u>

  Soit G = (V, ∑, R, S) une grammaire algébrique sous forme normale de Chomsky.

  Soit $S \Rightarrow_G^* w$ une dérivation de $w \in L(G)$.

  Si $|w| \geq 2^n$ alors l'arbre de dérivation est de hauteur $\geq n+1$.

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- <u>Exemple</u>

  Montrons que $L = \{ a^i b^i c^i \mid i \geq 0\}$ est non algébrique.

  Supposons que L est algébrique.

  D'après le lemme de la double étoile, il existe une constante k, dépendant de L, telle que :

  $\forall$ $z \in L$, $|z| \geq k$, z peut être décomposé en $z = uvwxy$ avec :

  (*i*) $|vwx| \leq k$

  (*ii*) $|v| + |x| > 0$

  (*iii*) $uv^n wx^n y \in L$, $\forall$ $n \geq 0$

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Exemple

Considérons la chaîne particulière $z_0 = a^k b^k c^k$.

On a bien $z_0 \in L$ et $|z_0| = 3k \geq k$.

Les décompositions de z_0 =uvwxy satisfaisant $|vwx| \leq k$ et $|v| + |x| > 0$ sont telles que :

– soit l'une des sous-chaînes v ou x contient plus d'un type de symbole, c-à-d de la forme a^+b^+ ou b^+c^+.

$\rightarrow uv^nwx^ny$ avec $n > 1$ contient un a après un b ou un b après un c.

(par exemple uv^2wx^2y = u aabb aabb w x x y, si v = aabb)

donc la chaîne uv^nwx^ny n'est plus de la forme $a^pb^pc^p$ avec $p \geq 0$,

donc $uv^nwx^ny \notin L$ pour $n > 1$.

– soit v et x sont des sous-chaînes de a^k ou de b^k ou de c^k.

Comme au plus une des chaînes v ou x est vide, toute chaîne de la forme uv^nwx^ny avec $n > 1$ est caractérisée par une augmentation de un (v = e ou x = e) ou deux ($v \neq e$ et $x \neq e$) des trois types de terminaux.

donc pour $n > 1$, la chaîne uv^nwx^ny est de la forme $a^pb^qc^r$ mais avec $p \neq q$ ou $q \neq r$.

donc $uv^nwx^ny \notin L$ pour $n > 1$.

Pour toutes les décompositions possibles de la chaîne z_0 il y a une contradiction.

Donc l'hypothèse est fausse.

$\Rightarrow$ L non algébrique.

Propriétés des langages algébriques

Lemme de l'étoile pour les langages algébriques

- Preuve de non algébricité

  Pour montrer qu'un langage est non algébrique, on peut utiliser :

  - le lemme de la double étoile,
  - les propriétés de stabilité de la classe des langages algébriques,
  - le théorème qui dit que l'intersection d'un langage algébrique et d'un langage rationnel est algébrique.

Problèmes indécidables pour les langages algébriques

- <u>Notion de problème indécidable</u>

  Une question est <u>décidable</u> s'il existe un <u>algorithme</u> (c'est-à-dire un processus <u>déterministe</u>) qui s'arrête avec une réponse (oui ou non) pour <u>chaque</u> entrée.

  Une question est <u>indécidable</u> si un tel algorithme n'existe pas.

Problèmes indécidables pour les langages algébriques

- Théorème

  Les questions suivantes sont décidables :

  – Étant donnés une grammaire algébrique G et un mot w, est-ce que $w \in L(G)$?

  – Étant donnée une grammaire algébrique G, est-ce que $L(G) = \varnothing$?

  Les questions suivantes sont indécidables :

  – *Soit G une grammaire algébrique. Est-ce que $L(G) = \Sigma^*$?*

  – *Soient G_1 et G_2 deux grammaires algébriques. Est-ce que $L(G_1) = L(G_2)$?*

  – *Soient M_1 et M_2 deux automates à pile. Est-ce que $L(M_1) = L(M_2)$?*

  – *Soit M un automate à pile. Trouver un automate à pile équivalent minimal en nombre d'états.*

À suivre …