Grammaires et Analyse Syntaxique - Cours 9 Propriétés des langages algébriques

Ralf Treinen

Université Paris Cité

IRIF Institut de Recherche en Informatique Fondamentale

treinen@irif.fr

21 mars 2024

Propriétés de clôture

- Nous avons vu en L2 que la classe des langages *rationnels* est clos sour presque toutes les opérations qu'on peut imaginer :
- concaténation, étoile de Kleene, union (évident car c'est dans la définition des expressions rationnelles),
- (c-à-d par exemple : Si L_1 et L_2 sont rationnels alors $L_1 \cup L_2$ l'est également.)
- ... mais aussi complément (car c'est le cas pour les langages *reconnaissables*, et rationnel = reconnaissable)
- ... et par conséquent aussi sous intersection !
- Est-ce pareil pour la classe des langages algébriques ?

Propriétés de clôture : union

- Les langages algébriques sont clos sous union : Si L_1 et L_2 sont algébriques, alors $L_1 \cup L_2$ est aussi algébrique.
- Preuve : Soient les grammaires

$$\begin{array}{ll} G_1 = (\Sigma_1, N_1, S_1, P_1) & \text{avec } L_1 = \mathcal{L}(G_1) \\ G_2 = (\Sigma_2, N_2, S_2, P_2) & \text{avec } L_2 = \mathcal{L}(G_1) \end{array}$$

- On suppose que N_1 et N_2 sont disjoints.
- Grammaire pour $L_1 \cup L_2$:

$$(\Sigma_1 \cup \Sigma_2, N_1 \cup N_2 \cup \{S_{new}\}, S_{new}, P_1 \cup P_2 \cup \{S_{new} \to S_1 \mid S_2\})$$

Propriétés de clôture : concaténation

- Les langages algébriques sont clos sous concaténation : Si L_1 et L_2 sont algébriques, alors $L_1 L_2$ est aussi algébrique.
- Preuve : Soient les grammaires

$$\begin{array}{ll} G_1 = (\Sigma_1, N_1, S_1, P_1) & \text{avec } L_1 = \mathcal{L}(G_1) \\ G_2 = (\Sigma_2, N_2, S_2, P_2) & \text{avec } L_2 = \mathcal{L}(G_1) \end{array}$$

- On suppose que N_1 et N_2 sont disjoints.
- Grammaire pour $L_1 L_2$:

$$(\Sigma_1 \cup \Sigma_2, N_1 \cup N_2 \cup \{S_{new}\}, S_{new}, P_1 \cup P_2 \cup \{S_{new} \to S_1\, S_2\})$$

Propriétés de clôture : étoile de Kleene

- Les langages algébriques sont clos sous itération (étoile de Kleene) : Si L est algébrique, alors L^* l'est également.
- Preuve : Soient la grammaire

$$G = (\Sigma, N, S, P) \quad \text{avec } L = \mathcal{L}(G)$$

- On suppose que $S_{new} \notin N$.
- Grammaire pour L^* :

$$(\Sigma, N \cup \{S_{new}\}, S_{new}, P \cup \{S_{new} \to \epsilon \mid S\, S_{new}\})$$

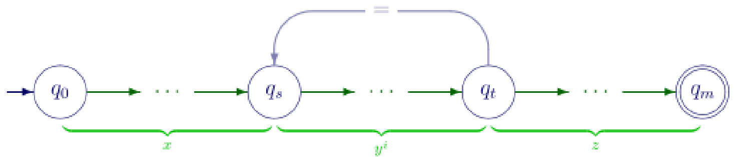
Montrer qu'un langage n'est pas algébrique

- Rappel : un langage L est *algébrique* s'il existe une grammaire algébrique G avec $L = \mathcal{L}(G)$.
- Donc pour montrer qu'un langage est algébrique il suffit de trouver une grammaire.
- Comment faire pour montrer qu'un langage n'est *pas* algébrique ?
- Il nous faut une méthode pour ça quand on veut exploiter les limites des grammaires.

Souvenirs du cours AAL3 : le lemme d'itération

- On a vu en AAL3 une méthode pour montrer qu'un langage n'est pas rationnel : le *lemme d'itération* (ou lemme d'étoile, *pumping lemma*) :
- Pour tout langage régulier L il existe un $N \geq 1$ tel que pour tout $u \in L$ avec $|u| \geq N$ existe une décomposition $u = xyz$ avec
  1. $y \neq \epsilon$
  2. $|xy| \leq N$
  3. $xy^iz \in L$ pour tout $i \geq 0$
- Intuitivement : la capacité de mémorisation d'un automate est limitée par son nombre d'états.

Illustration du lemme d'itération des langages réguliers

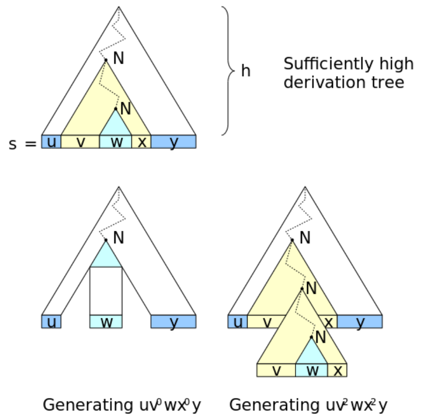


Source : Jochen Burghardt/wikipedia

Des langages réguliers aux langages algébriques

- Le lemme d'itération des langages réguliers vient des automates : Si un mot d'entrée est suffisamment long alors il y un état de l'automate qui se répète. On peut alors itérer le mot lu entre ces deux occurrences du même état.
- Pour les langages algébriques on peut obtenir un lemme similaire, basé sur les arbres de dérivation : si un arbre de dérivation est suffisamment profond alors il y a non-terminal de la grammaire qui se répète sur une branche de cet arbre.
- On peut alors itérer le mot produit "entre ces deux occurrences du même non-terminal" (voir le dessin).

Idée de la preuve du lemme d'itération



Source : Jochen Burghardt/wikipedia

Un lemme d'itération pour les langages algébriques

Lemme d'itération pour les langages algébriques

Soit L un langage algébrique. Alors il existe un $p \geq 1$ tel que tout mot $s \in L$ avec $|s| \geq p$ peut être décomposé en $s = uvwxy$ avec :

- $|vwx| \leq p$
- $|vx| \geq 1$
- $uv^iwx^iy \in L$ pour tout $i \geq 0$

Négation de la propriété d'itération

- Donc on peut montrer qu'un langage L n'est *pas* algébrique si on montre que L satisfait la *négation* de la propriété d'itération.
- C'est-à-dire il faut montrer :
  - $\forall p$ avec $p \geq 1$:
  - $\exists s \in L$ avec $|s| \geq p$
  - $\forall u, v, w, x, y$ avec $s = uvwxy$, $|vwx| \leq p$, $|vx| \geq 1$
  - $\exists i$ tel que $uv^iwx^iy \notin L$
- Comment faire une telle preuve ?

Utiliser le lemme d'itération

- On peut voir une telle formule logique avec une alternance de quantificateurs $\forall$ et $\exists$ comme un jeu entre deux joueurs : Existentiel et Universel.
- Existentiel cherche à montrer que la formule est vraie, Universel cherche à l'empêcher.
- Les deux choisissent des valeurs de variables : Existentiel pour les variables existentielles, Universel pour les variables universelles.
- Les joueurs peuvent prendre en considération les choix précédents de leur adversaire.
- La formule est vrai si Existentiel a une *stratégie gagnante*.

Application : un langage qui n'est pas algébrique

- Nous allons montrer : $L = \{a^n b^n c^n \mid n \geq 0\}$ n'est pas algébrique.
- L'opposant choisi une valeur $p \geq 1$.
- Nous choisissons $s = a^p b^p c^p$.
- L'opposant choisi une décomposition $uvwxy = a^p b^p c^p$ avec $|vwx| \leq p$ et $|vx| \geq 1$.
- Il est maintenant à nous de choisir i, et de démontrer que $uv^i wx^i y \notin L$. On considère les cas différents ou dans s se trouve la partie vwx :

Application : un langage qui n'est pas algébrique

- Si $vwx \in a^+$: il existe k, l avec $v = a^k$ et $x = a^l$ et $k + l > 0$. Nous choisissons $i = 0$, on a alors $uwy = a^{p-k-l} b^p c^p \notin L$ car $p - k - l \neq p$.
- Si $vwx \in b^+$: pareil
- Si $vwx \in c^+$: pareil
- Si $vwx \in a^+ b^+$: on a que v ou x contiennent au moins un a ou un b, mais pas de c. Nous choisissons $i = 0$, on a alors $|uwy|_a < |uvwxy|_a$ ou $|uwy|_b < |uvwxy|_b$, mais $|uwy|_c = |uvwxy|_c$. Donc $uwy \notin L$.
- Si $vwx \in b^+ c^+$: pareil
- Si $vwx \in a^+ b^+ c^+$: pas possible car $|vwx| \leq p$.

Propriétés de clôture : intersection ?

- Les langages algébriques ne sont *pas* clos sous intersection : il existe deux langages algébriques L_1 et L_2 tel que $L_1 \cap L_2$ n'est pas algébrique.
- $L_1 = \{a^n b^n c^m \mid n, m \geq 0\}$
- L_1 est algébrique car concaténation de deux langages algébriques :
  - $\{a^n b^n \mid n \geq 0\}$, grammaire $S \to aSb \mid \epsilon$;
  - c^*, c'est même un langage rationnel.
- $L_2 = \{a^m b^n c^n \mid n, m \geq 0\}$, algébrique pour la même raison.
- Or, $L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 0\}$ qui n'est pas algébrique.

Propriétés de clôture : complément ?

- Les langages algébriques ne sont *pas* clos sous complément.
- Preuve :
  - Supposons pour l'absurde que le complément $\overline{L}$ de tout langage algébrique L est aussi algébrique.
  - Il en suit que les langages algébriques sont aussi clos sous intersection, car
  $$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$$
  - Contradiction avec le transparent précédent.