I53: Théorie des langages et compilation
Contrôle terminal - session 1
Licence Informatique
Correction

2019-2020 (semestre 1) - Durée : 3h00

- Tous les documents, calculatrices et appareils de communication sont interdits.
- Le barême est donné à titre indicatif
- Toutes les réponses doivent être justifiées

Exercice 1. **Questions de cours** (2 pts)

1. Rappeler le rôle des phases d'analyse lexicale et syntaxique lors de la compilation d'un programme.

   Solution

   - Le rôle de l'analyse lexicale est de décomposer le flux de donnée en entrée en unités lexicales.
   - Le rôle de l'analyse syntaxique est de créer une représentation du flots d'unités lexicales reflétant la structure grammaticale ou hierarchique du lanagage.

2. Quel est le rôle de la table de symboles ?

   Solution

   - Le rôle de la table de symboles est de centraliser les informations ratachées aux identificateurs du programme source et peut être utilisée et mise à jour lors des différentes phases de la compilation.

3. Quel est l'intérêt de produire du code intermédiaire (type code à trois adresses) avant de produire du code en langage cible ?

   Solution

   - L'intérêt principal est de faliciter la conception des compilateurs d'un langage donné vers plusieurs architectures cibles. La partie frontale reste la même et seul le programme de production de code nécessite d'être adapté.

Exercice 2. **Expressions régulières et automates** (5 pts)

1. Donner des définitions régulières des langages suivants:

    (a) les chaînes sur $\{a, b\}$ telles que chaque a est suivi d'exactement 1 ou 3 b;

    (b) les chaînes sur $\{a, b\}$ telles que le nombre de a suivi d'un b est pair.

    Solution

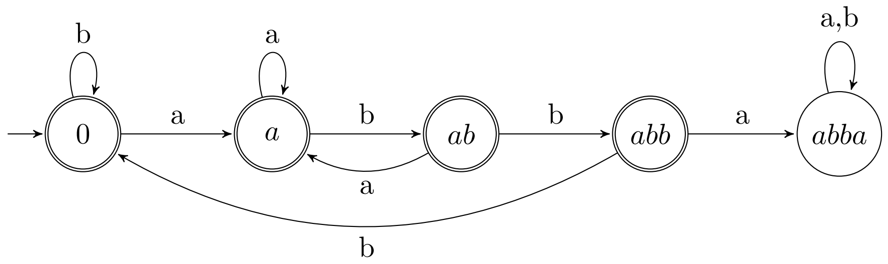
    (a) $b^*(a(b + bbb))^*$

    (b) $b^*(aa^*bb^* + aa^*bb^*)^*a^*$

2. Dessiner les automates reconnaissant les langages suivants:

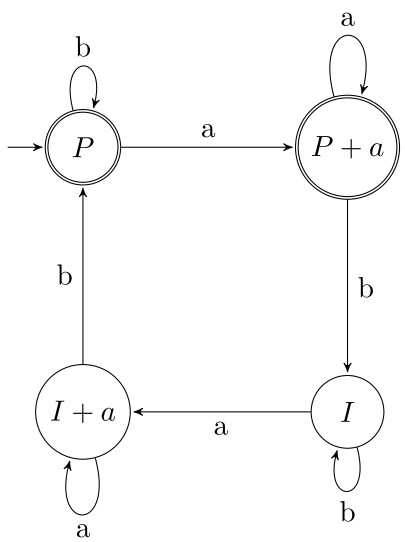
    (a) les chaînes sur $\{a, b\}$ telles que le motif $abba$ n'apparait pas;

    Solution

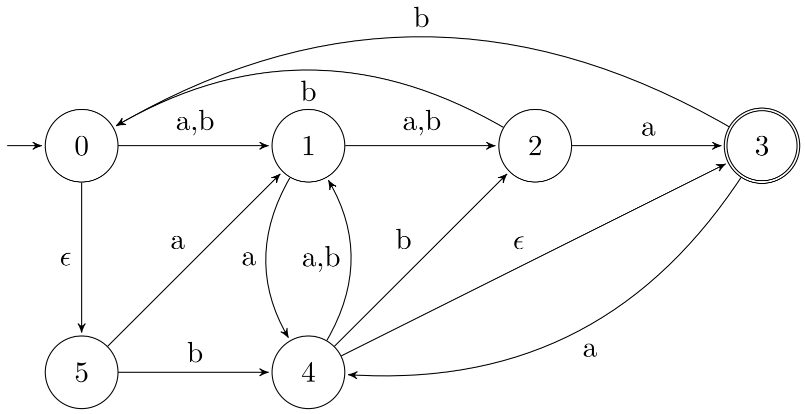


    (b) les chaînes sur $\{a, b\}$ telles que le nombre de a suivi d'un b est pair.

    Solution



3. Déterminiser l'automate suivant:



Solution Utiliser le programme vu en TP !

Exercice 3. racine carrée d'un langage (8 pts)

Soit w un mot sur un alphabet Σ, on rappelle que $|w|$ représente la longueur de w. Pour tout symbole σ de l'alphabet on définit $|w|_\sigma$ comme le nombres d'occurences de σ dans w (par exemple, pour $w = abbabab$, $|w|_a = 3, |w|_b = 4$).

Soit A un automate fini. On rappelle qu'un état q de l'automate est *accessible* s'il existe un chemin depuis l'état initial jusqu'à q.

Soit L un langage sur un alphabet Σ, on appelle racine carrée de L le langage

$$\sqrt{L} = \{w \in \Sigma^* | ww \in L\}.$$

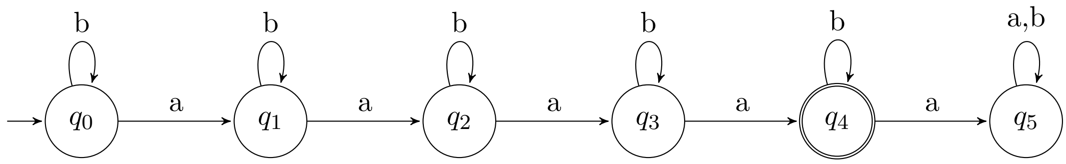
Par exemple, pour $L = \{ab, aabb, aba, bb, baabbaab\}$, $\sqrt{L} = \{b, baab\}$.

1. Calculer la racine carrée des langages suivants:
   (a) $L_1 = \{abba, baba, abababa, abbabb, abbbababbbab, bbabbabbabba\}$;
   (b) $L_2 = \{ab, abab, ababab, abababab, \dots\}$;
   (c) $L_3 = \{w \in \{a,b\}^* \text{ tq } |w|_a = n\}$ (pour un entier $n \geq 0$ fixé).

   Solution

   (a) $\sqrt{L_1} = \{ab, ba, abb, abbbab, bbabba\}$
   (b) $\sqrt{L_2} = \{ab, abab, ababab, \dots\} = L_2$
   (c) $\sqrt{L_3} = \{w \in \{a,b\}^* \text{ tq } |w|_a = n/2\}$ si n est pair et $\sqrt{L_3} = \emptyset$ sinon.

2. Soit B l'automate suivant:



   Décrire le langage $L(B)$.

   Solution $L(B)$ est le langage des mots sur $\{a,b\}$ contenant exactement 4 a.

3. On définit un parcours simultané de l'automate comme suit: étant donné un couple d'états (s, s') et un symbole σ (ici a ou b) le couple d'états suivant est simplement $(s.\sigma, s'.\sigma)$. Par exemple $(q_0, q_1) \xrightarrow{a} (q_1, q_2)$ et $(q_4, q_5) \xrightarrow{a} (q_5, q_5)$.

   Effectuer un parcours simultané des mots $baab$ et $babaa$ à partir du couple d'états (q_0, q_2).

   Solution

$$(q_0, q_2) \xrightarrow{b} (q_0, q_2) \xrightarrow{a} (q_1, q_3) \xrightarrow{a} (q_2, q_4) \xrightarrow{b} (q_2, q_4)$$

$$(q_0, q_2) \xrightarrow{b} (q_0, q_2) \xrightarrow{a} (q_1, q_3) \xrightarrow{b} (q_1, q_3) \xrightarrow{a} (q_2, q_4) \xrightarrow{a} (q_3, q_5)$$

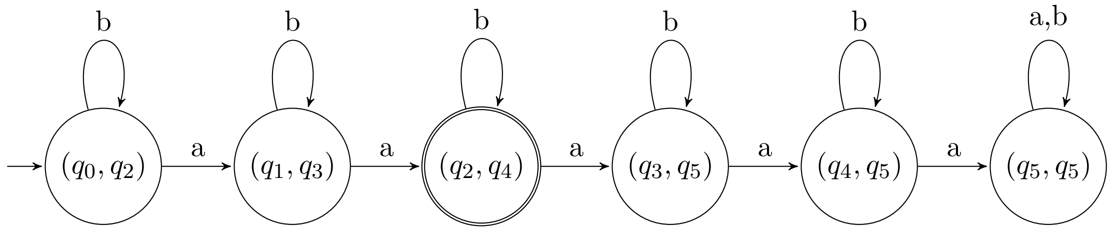
4. Soit $A = \{Q, \Sigma, q_0, F, \delta\}$ un AFD. On veut construire formellement un automate permettant de parcourir A simultanément à partir de q_0 et d'un autre état.

   Soit q un état de Q fixé. On définit ce nouvel automate $A_q = \{Q_q, \Sigma, I_q, F_q, \delta_q\}$ comme suit:

- $Q_q = Q \times Q = \{(s, s')|s \in Q, s' \in Q\}$
- $I_q = (q_0, q)$
- $F_q = \{(q, q_F)|q_F \in F\}$
- $\forall (s, s') \in Q_q, \delta_q((s, s'), \sigma) = (\delta(s, \sigma), \delta(s', \sigma)) = (s.\sigma, s'.\sigma)$

En reprenant l'automate B précédent, dessiner B_{q_2} en ne considérant que les états accessibles (par exemple (q_0, q_1) n'est pas accessible). Décrire le langage $L(B_{q_2})$ en fonction de B.

Solution



$L(B_{q_2})$ est le langage des mots contenant exactement 2 a, c'est-à-dire $\sqrt{L(B)}$.

5. Soit w un mot. Montrer que

$$w \in L(A_q) \Rightarrow w \in \sqrt{L(A)}.$$

Solution Par définition de $L(A_q)$ on a $(q_0.w, q.w) = (q, q_F)$ avec q_F un état final de A. On a ainsi $q_0.(ww) = (q_0.w).w = q.w = q_F$ et donc $ww \in L(A)$ (autrement dit $w \in \sqrt{L(A)}$.

6. Réciproquement, montrer que

$$w \in \sqrt{L(A)} \Rightarrow \exists q \in Q \text{ tq } w \in L(A_q).$$

Solution

On a $ww \in L(A)$ donc $q_0.(ww) = q_F$ avec q_F un êtat final de A. Posons $q = (q_0.w)$ et considérons l'automate A_q comme défini précédemment.

On a bien $(q_0.w, q.w) = (q, q_F)$ avec q_F un êtat final de A et donc $w \in L(A_q)$.

7. En déduire que la racine carrée d'un langage rationnel L est aussi un langage rationnel.

Solution Soit L un langage rationnel et $A = (Q, \Sigma, q_0, F, \delta)$ un AFD acceptant L. D'après les questions 5 et 6, on a l'équivalence suivante:

$$w \in \sqrt{L(A)} \Leftrightarrow \exists q \in Q \text{ tq } w \in L(A_q).$$

Considérons maintenant $L' = \bigcup_{q \in Q} L(A_q)$. L' est un langage rationnel en tant qu'union de langages rationnels. De plus

$$w \in L' \Leftrightarrow \exists q \in Q \text{ tq } w \in L(A_q) \Leftrightarrow w \in \sqrt{L(A)}.$$

On a donc $L' = \sqrt{L(A)} = SqrtL$ qui est donc bien un langage rationnel.

Exercice 4. Grammaire des expressions régulières (5 pts)

On considère le langage des expressions régulières sur un alphabet quelconque défini par la grammaire $G = (T = \{symb, +, *, (,)\}, N = \{E\}, S = E, \mathcal{P})$ où le terminal *symb* représente n'importe quel symbole de l'alphabet et dont les productions sont les suivantes:

$$E \rightarrow E + E \mid EE \mid E* \mid (E) \mid symb$$

1. Montrer à l'aide de la chaîne $a + a(a + b)$ que la grammaire est ambiguë.
2. On considère maintenant la grammaire équivalente G' suivante:

$$\begin{array}{lcl} E & \rightarrow & E + T \mid T \\ T & \rightarrow & TK \mid K \\ K & \rightarrow & K* \mid F \\ F & \rightarrow & (E) \mid symb \end{array}$$

3. Donner une grammaire équivalente G'' sans récurisivité à gauche.
4. Calculer $Premier(A)$ et $Suivant(A)$ pour tous les non-terminaux A de G''.
5. En déduire une table d'analyse syntaxique pour G''.

Solution Voir TD.