

**Cursus:** M1, computer-science  
**Code UE:** JEIN8602  
**Subject:** Formal languages theory  
**Date:** March 4th 2017  
**Duration:** 3H  
 Documents: authorized  
 Lectures by: Mr Géraud Sénizergues

The exercises are *independent* one from each other. Thus each mathematical symbol ( $G, L, \dots$ ) has a single definite meaning *inside* each exercise, but might have different meanings from an exercise to another.

It is *not* required to solve *all* the exercises. Every correct solution to an exercise will give (around) 4 points. A question marked with a (\*) is difficult. One can admit the result of such a question and, nevertheless, *use* it in subsequent questions.

**Exercise 1** [/4] We consider the finite automaton  $\mathcal{A}$  described on figure 1. Note that 0 is

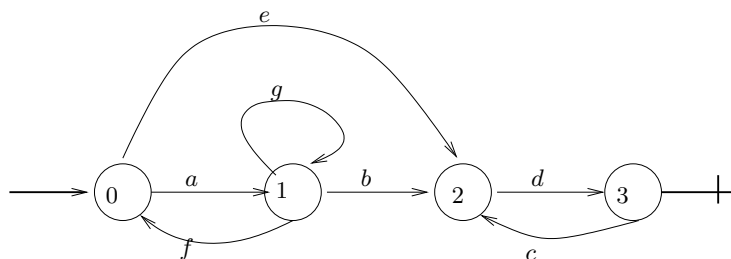


Figure 1: finite automaton  $\mathcal{A}$

the only initial state and 3 is the only final state.

0- Describe accepting computations of  $\mathcal{A}$  over the words:

$agbd, ed, abdc, afed.$

1- Construct a regular expression for the language  $L_{\mathcal{A}}$  recognized by the automaton  $\mathcal{A}$ . Explain the successive steps of your construction.

We consider the monoid homomorphism  $h : \{a, b, c, d, e, f, g\}^* \rightarrow \{a, b, c\}^*$  defined by:

$$h(a) = a, \quad h(b) = b, \quad h(c) = ba, \quad h(d) = c, \quad h(e) = aa, \quad h(f) = ba, \quad h(g) = a.$$

2- Give a regular expression for the language  $h(L_{\mathcal{A}})$ .

3- Construct a finite automaton recognizing the language  $h(L_{\mathcal{A}})$ .

**Exercise 2** [/4] Let us consider the regular expression:

$$e := (ab^*c)^*ab(a \cup b)^*$$

Construct, by Glushkov's method, a finite automaton recognizing  $L_e$ .

**Exercise 3** [/6]

1- Let us consider the language  $L_2 := \{a, b\}^*abb\{a, b\}^* \subseteq \{a, b\}^*$ . Construct a finite automaton, over the alphabet  $X = \{a, b\}$ , that recognizes  $L_2$ .

2- Compute the *minimal complete deterministic* automaton of  $L_2$ .

3- Compute a regular expression for the language  $X^* - L_2$  (the *complement* of  $L_2$ ).

4\*- For every natural integer  $n \geq 1$  we define

$$L_n := \{a, b\}^*ab^n\{a, b\}^*.$$

How many states does possess the minimal complete deterministic automaton of  $L_n$ ?

5- Let us consider the languages:

$$L := \bigcup_{n \geq 1} L_n; \quad M_n := L_n \cdot L_{n+1} \text{ (for } n \geq 1); \quad M := \bigcup_{n \geq 1} M_n$$

Is  $L$  regular ?

Is it true that, for every  $n \geq 1$ ,  $M_n$  is regular ?

Is  $M$  regular ?

**Exercise 4** [/4]

1- For each of the following languages over the terminal alphabet  $\{a, b, c, d\}$ , construct a context-free grammar that generates the language  $L_i$ :

$$L_1 := \{(ab)^n \mid n \geq 0\}$$

$$L_2 := \{(ab)^n c^n \mid n \geq 0\}$$

$$L_3 := \{(ab)^n c^m \mid n \geq m \geq 0\}$$

$$L_4 := \{(ab)^p c^q d^r \mid q \geq 0, r \geq 0, p = q + r\}$$

$$L_5 := \{(ab)^n c^m \mid n \geq 0, m \geq 0, n \neq m\}$$

2- Construct a *non-ambiguous* context-free grammar generating the language  $L_5$ .

**Exercise 5** [/4] We consider the context-free grammar  $G := (A, N, R)$  where  $A = \{a, b, c\}$ ,  $N = \{S_1, S_2, S_3, S_4, S_5\}$  and  $R$  consists of the following 12 rules:

$$\begin{array}{lll} S_1 \rightarrow aS_1S_1 & S_1 \rightarrow bS_4S_1 & S_1 \rightarrow S_3c \\ S_2 \rightarrow aS_2 & S_2 \rightarrow aS_5a & S_3 \rightarrow S_3S_1 \\ S_3 \rightarrow aS_4 & S_3 \rightarrow S_1S_3S_1 & S_4 \rightarrow a \\ S_4 \rightarrow S_1S_4 & S_5 \rightarrow cS_5 & S_5 \rightarrow aS_5S_2 \end{array}$$

The start symbol of  $G$  is  $S_1$ .

- 1- What are the *productive* non-terminals of  $G$  ?
- 2- What are the *useful* non-terminals of  $G$  ?
- 3- Transform the grammar  $G$  into an equivalent grammar  $G'$  where every non-terminal is productive and useful.
- 4- Is the language  $L(G, S_1)$  empty ?
- 5- Is the language  $L(G, S_1)$  infinite ?

**Exercise 6** [/5] We consider the context-free grammar  $G := (A, N, R)$  where  $A = \{a, b, c, d\}$ ,  $N = \{S_0, S_1, S_2, S_3\}$  and  $R$  consists of the following rules:

- r1:**  $S_0 \rightarrow aS_1$
- r2:**  $S_0 \rightarrow aS_2$
- r3:**  $S_1 \rightarrow bS_2$
- r4:**  $S_1 \rightarrow cS_3$
- r5:**  $S_2 \rightarrow bS_1$
- r6:**  $S_2 \rightarrow dS_3$
- r7:**  $S_3 \rightarrow \varepsilon$

The start symbol of  $G$  is  $S_0$ .

- 1- Show that  $bbbcb \in L(G, S_1)$ ,  $bbbbbd \in L(G, S_1)$  . Give a derivation from  $S_1$  to  $bbbcb$ .
- 2- Show that  $bbbbbd \in L(G, S_2)$ ,  $bbbbbc \in L(G, S_2)$  . Give a derivation from  $S_2$  to  $bbbbbd$ .
- 3- Show that  $abbbbd \in L(G, S_0)$ .

Give a derivation-*tree* for  $abbbbd$ .

(The student can choose, among the various notions of derivation-tree, syntax-tree, abstract syntax-tree, etc... his/her favorite notion and compute the corresponding tree).

4- Describe, the languages  $L(G, S_1)$ ,  $L(G, S_2)$ .

Is it true that  $L(G, S_1) \cap L(G, S_2) = \emptyset$  ?

4- Is the grammar  $G$  ambiguous ? Is the grammar  $G$  simple ?

5- Could you construct a *simple* c.f. grammar  $G'$  with starting symbol  $S'$ , such that  $L(G', S') = L(G, S_0)$  ?

We consider the context-free grammar  $H := (A_H, N_H, R_H)$  where  $A_H = \{a, b, c, d, \#\}$ ,  $N_H = \{T_0, U_1, U_2, T_1, T_2, T_3\}$  and  $R_H$  consists of the following rules:

- r1:**  $T_0 \rightarrow U_1a$
- r2:**  $T_0 \rightarrow U_2a$
- r3:**  $U_1 \rightarrow aT_1$
- r4:**  $U_2 \rightarrow aT_2$
- r5:**  $T_1 \rightarrow bT_2b$
- r6:**  $T_1 \rightarrow cT_3c$
- r7:**  $T_2 \rightarrow bT_1b$
- r8:**  $T_2 \rightarrow dT_3d$
- r9:**  $T_3 \rightarrow \#$

The start symbol of  $H$  is  $T_0$ .

6- Is  $H$  ambiguous ? simple ?

7- Could you construct a *simple* c.f. grammar  $H'$  with starting symbol  $T'$ , such that  $L(H', T') = L(H, T_0)$  ?

Hint: Remove non-terminals  $U_1, U_2$  (by a suitable transformation that preserves the generated language). Then, let be inspired by the analogy with question 5.