Department of Information and Computing Sciences
Utrecht University

INFOB3TC – Solutions for Exam 2

Johan Jeuring

Wednesday, 1 February 2017, 11:00–13:00

Please keep in mind that there are often many possible solutions and that these example solutions may contain mistakes.

Multiple-choice questions

In this series of 10 multiple-choice question, you get:

- 5 points for each correct answer,
- 1 point if you do not answer the question,
- and 0 points for a wrong answer.

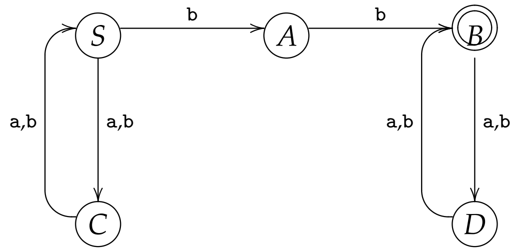
Answer these questions with *one of* a, b, c, or d.

1 (5 points)**.** Consider the following regular language:

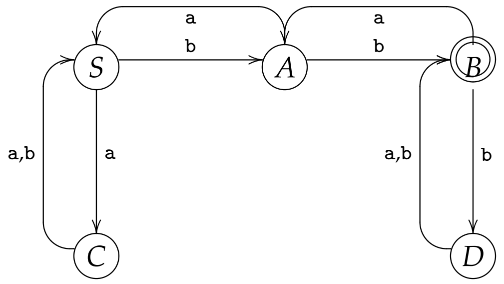
$$L \;=\; \{x \mid x \in \{\texttt{a},\texttt{b}\}^*, \text{length } x \text{ is even}, \texttt{bb} \text{ is a substring of } x\}$$

Which of the following automata, with start state S, generates L?

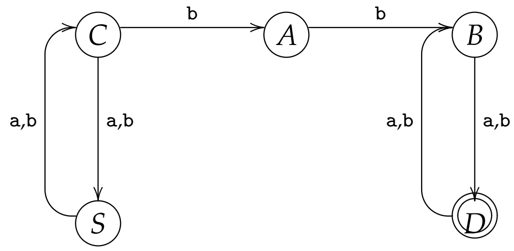
a)



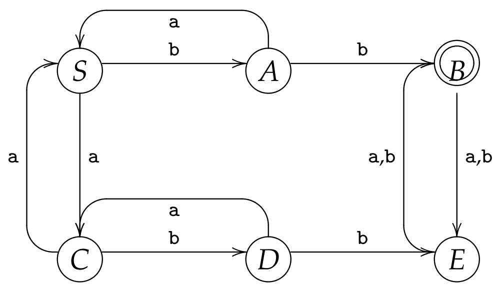
b)



c)

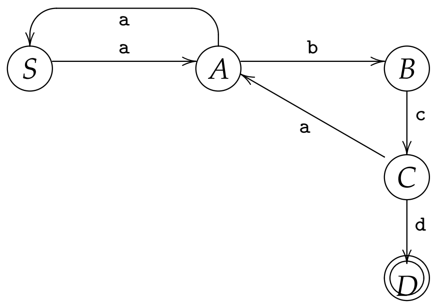


d)



Solution 1. d). a), b) and c) cannot produce abba.

2 (5 points). Consider the following nondeterministic finite state automaton with starting state S.



Which of the following regular expressions generates the same language as this automaton?

a) $\texttt{a}(\texttt{aa} + \texttt{bca})*\texttt{bcd}$.

b) $\texttt{a}((\texttt{aa})* + (\texttt{bca})*)\texttt{bcd}$.

c) $\texttt{aa}*(\texttt{bca})*\texttt{bcd}$.

d) $\texttt{a}(\texttt{aa})*(\texttt{bca})*\texttt{bcd}$.

•

Solution 2. a). ○

3 (5 points). Consider the regular language

$$\{x \in \{\texttt{a}, \texttt{b}, \texttt{c}\}^* \mid \text{the number of a's plus the number of b's is even}\}$$

Which of the following regular expressions does *not* describe this language?

a) $(\texttt{c}^*(\texttt{a} + \texttt{b})\texttt{c}^*(\texttt{a} + \texttt{b})\texttt{c}^*)^* + \texttt{c}^*$.

b) $\texttt{c}^*((\texttt{a} + \texttt{b})\texttt{c}^*(\texttt{a} + \texttt{b}))^*\texttt{c}^*$.

c) $(\texttt{c}^*(\texttt{a} + \texttt{b})\texttt{c}^*(\texttt{a} + \texttt{b}))^*\texttt{c}^*$.

d) $\texttt{c}^*((\texttt{a} + \texttt{b})\texttt{c}^*(\texttt{a} + \texttt{b})\texttt{c}^*)^*$.

•

Solution 3. b). ○

4 (5 points)**.** Consider the following two languages on the terminal symbols x and y:

$$\begin{aligned} L_1 &= \{\mathtt{x}^i\mathtt{y}^j \mid j \geqslant i \geqslant 10\} \\ L_2 &= \{\mathtt{x}^i\mathtt{y}^j \mid 10 \geqslant i \geqslant j\} \end{aligned}$$

a) Only L_2 is regular.

b) Only L_1 is regular.

c) None of L_1 and L_2 are regular.

d) L_1 and L_2 are both regular.

Solution 4. a).

5 (5 points)**.** I want to show that the language L:

$$\{x \mid x \in \{\mathtt{a},\mathtt{b}\}^*, nr\ \mathtt{a}\ x < nr\ \mathtt{b}\ x\}$$

where $nr\ \mathtt{c}\ y$ is the number of occurrences of c in y, is not regular. Then I have to show that for all natural numbers n there exist $x,\ y,\ z$ such that $xyz \in L$ en $|y| \geqslant n$, such that ... Which of the following choices for $x,\ y,\ z$ ensure that I can easily complete the proof?

a) $z = \epsilon, y = \mathtt{a}^n\mathtt{b}^n, x = \mathtt{b}^n$.

b) $x = \epsilon, y = \mathtt{b}^{2n}, z = \mathtt{a}^n$.

c) $x = \mathtt{a}^n, y = \mathtt{b}^n, z = \mathtt{b}^n$.

d) $x = \mathtt{b}^{2n}, y = \mathtt{a}^n, z = \epsilon$.

Solution 5. d).

6 (5 points)**.** Consider the following two languages:

$$L_1 = \{0^n 1^m 0^n 1^m \mid n, m > 0\}$$
$$L_2 = \{0^n 1^m 0^m 1^n \mid n, m > 0\}$$

Are L_1 and L_2 context-free?

a) Both L_1 and L_2 are context-free.

b) None of L_1 and L_2 are context-free.

c) Only L_1 is context-free.

d) Only L_2 is context-free.

•

Solution 6. d) ◦

7 (5 points)**.** For which nonterminals N of the following grammar

$$S \rightarrow A\mathtt{a}C \mid B\mathtt{d}$$
$$A \rightarrow BC$$
$$B \rightarrow \mathtt{b}B \mid C$$
$$C \rightarrow \mathtt{acc}S \mid [\,]$$

does `first` N contain the terminal `b`?

a) $\{B\}$.

b) $\{A, B, S\}$.

c) $\{B, C\}$.

d) $\{A, B\}$.

•

Solution 7. b). ◦

8 (5 points). For which nonterminals N of the following grammar

$$
\begin{aligned}
S &\rightarrow A\texttt{a}C \mid B\texttt{d} \\
A &\rightarrow BC \\
B &\rightarrow \texttt{b}B \mid C \\
C &\rightarrow \texttt{acc}S \mid \epsilon
\end{aligned}
$$

is `d` element of the `follow` N set?

a) $\{B\}$.

b) $\{S, A, B, C\}$.

c) $\{B, C\}$.

d) $\{S, B, C\}$.

Solution 8. d).

9 (5 points)**.** Consider the following two grammars:

I

$$
\begin{array}{rcl}
W & \rightarrow & \texttt{c} \mid \texttt{b} \\
S & \rightarrow & T\texttt{a} \mid VT\texttt{a} \\
T & \rightarrow & ST \mid W \mid \texttt{a} \\
V & \rightarrow & \texttt{b}
\end{array}
$$

II

$$
\begin{array}{rcl}
S & \rightarrow & BC\texttt{d} \mid \epsilon \\
A & \rightarrow & A\texttt{a}S\texttt{b} \mid S\texttt{b}C \mid \epsilon \\
B & \rightarrow & \texttt{b} \mid \epsilon \\
C & \rightarrow & \texttt{c} \mid B
\end{array}
$$

Are these grammars LL(1)?

a) None of the two grammars is LL(1).

b) Both grammar I and grammar II are LL(1).

c) Only grammar I is LL(1).

d) Only grammar II is LL(1).

Solution 9. a).

10 (5 points)**.** Construct the LR(0) automaton for the following grammar.

$$\begin{aligned} S' &\rightarrow S\$ \\ S &\rightarrow X\texttt{x} \mid Y\texttt{y} \\ X &\rightarrow \texttt{x} \\ Y &\rightarrow \texttt{y} \end{aligned}$$

Which of the following statements is true?

a) The automaton has no conflicts.

b) The automaton has both a shift/reduce and a reduce/reduce conflict.

c) The automaton contains a shift/reduce conflict.

d) The automaton contains a reduce/reduce conflict.

•

Solution 10. a). ○

11 (5 points)**.** The different LR classes categorize grammars: LR(0) ⊂ SLR(1) ⊂ LALR(1) ⊂ LR(1). What is the smallest set of which the following grammar is a member?

$$\begin{aligned} S &\rightarrow E\$ \\ E &\rightarrow A\texttt{a}B \mid B \\ A &\rightarrow \texttt{b}B \mid \texttt{c} \\ B &\rightarrow A \end{aligned}$$

a) LR(1)

b) LALR(1)

c) SLR(1)

d) LR(0)

•

Solution 11. b). ○

Open answer questions

12 (15 points)**.** Consider the language P_{ab} given by the following context-free grammar with start symbol P:

$$P \rightarrow \mathtt{a}P\mathtt{a} \mid \mathtt{b}P\mathtt{b} \mid \mathtt{a} \mid \mathtt{b} \mid \epsilon$$

Prove that the language P_{ab} is not regular, using the pumping lemma for regular languages. •

Solution 12. The language P_{ab} is not regular. To prove it, we assume it is regular and find a contradiction using the pumping lemma.

For any n,

let $x = \varepsilon$, $y = \mathtt{a}^n$, $z = \mathtt{ba}^n$. Then $xyz = \mathtt{a}^n\mathtt{ba}^n \in L(P)$, and $|y| \geqslant n$.

From the pumping lemma, we know there must be a loop in y, i.e. $y = uvw$ with $q = |v| > 0$ such that $xuv^iwz \in L(P)$ for all $i \in \mathbb{N}$.

Let $i = 2$. We expect $xuv^2wz \in L(P)$. If $u = a^s$, $v = a^q$, $w = a^t$, then we get $\mathtt{a}^{s+2\,q+t}\mathtt{ba}^n \in L(P)$. But this word is not in L, since $q > 0$, and hence $s + 2\,q + t > n$. Therefore, P_{ab} is not regular. ○

13 (15 points)**.** Consider the following grammar:

$$
\begin{aligned}
&S \to E\{P\} \mid \varepsilon \\
&P \to V{=}S \mid \varepsilon \\
&V \to \texttt{a} \mid \texttt{b} \mid \texttt{c} \\
&E \to\ ! \mid ?D \\
&D \to PS
\end{aligned}
$$

To use this grammar in an LL(1) parser, we need to determine several properties of this grammar. Fill out the table below by computing the values in the columns for the appropriate rows. Use *True* and *False* for property values and set notation for everything else.

•

Solution 13.

NT	Production	*empty*	*emptyRhs*	*first*	*firstRhs*	*follow*	*lookAhead*
S		True		$\{!, ?\}$		$\{\{, \}, !, ?\}$	
	$S \to E\{P\}$		False		$\{!, ?\}$		$\{!, ?\}$
	$S \to \varepsilon$		True		$\{\}$		$\{\{, \}, !, ?\}$
P		True		$\{\texttt{a}, \texttt{b}, \texttt{c}\}$		$\{\{, \}, !, ?\}$	
	$P \to V{=}S$		False		$\{\texttt{a}, \texttt{b}, \texttt{c}\}$		$\{\texttt{a}, \texttt{b}, \texttt{c}\}$
	$P \to \varepsilon$		True		$\{\}$		$\{\{, \}, !, ?\}$
V		False		$\{\texttt{a}, \texttt{b}, \texttt{c}\}$		$\{=\}$	
	$V \to \texttt{a}$		False		$\{\texttt{a}\}$		$\{\texttt{a}\}$
	$V \to \texttt{b}$		False		$\{\texttt{b}\}$		$\{\texttt{b}\}$
	$V \to \texttt{c}$		False		$\{\texttt{c}\}$		$\{\texttt{c}\}$
E		False		$\{!, ?\}$		$\{\{\}$	
	$E \to\ !$		False		$\{!\}$		$\{!\}$
	$E \to\ ?D$		False		$\{?\}$		$\{?\}$
D		True		$\{\texttt{a}, \texttt{b}, \texttt{c}, !, ?\}$		$\{\{\}$	
	$D \to PS$		True		$\{\texttt{a}, \texttt{b}, \texttt{c}, !, ?\}$		$\{\{, \texttt{a}, \texttt{b}, \texttt{c}, !, ?\}$

Marking
One error in a column: -1
More than one error: -2.5
I didn't check the Rhs columns carefully, their function is to help fill out the other columns

◦

14 (15 points)**.** Consider the context-free grammar:

$$S \to AS$$
$$S \to \texttt{b}$$
$$A \to SA$$
$$A \to \texttt{a}$$

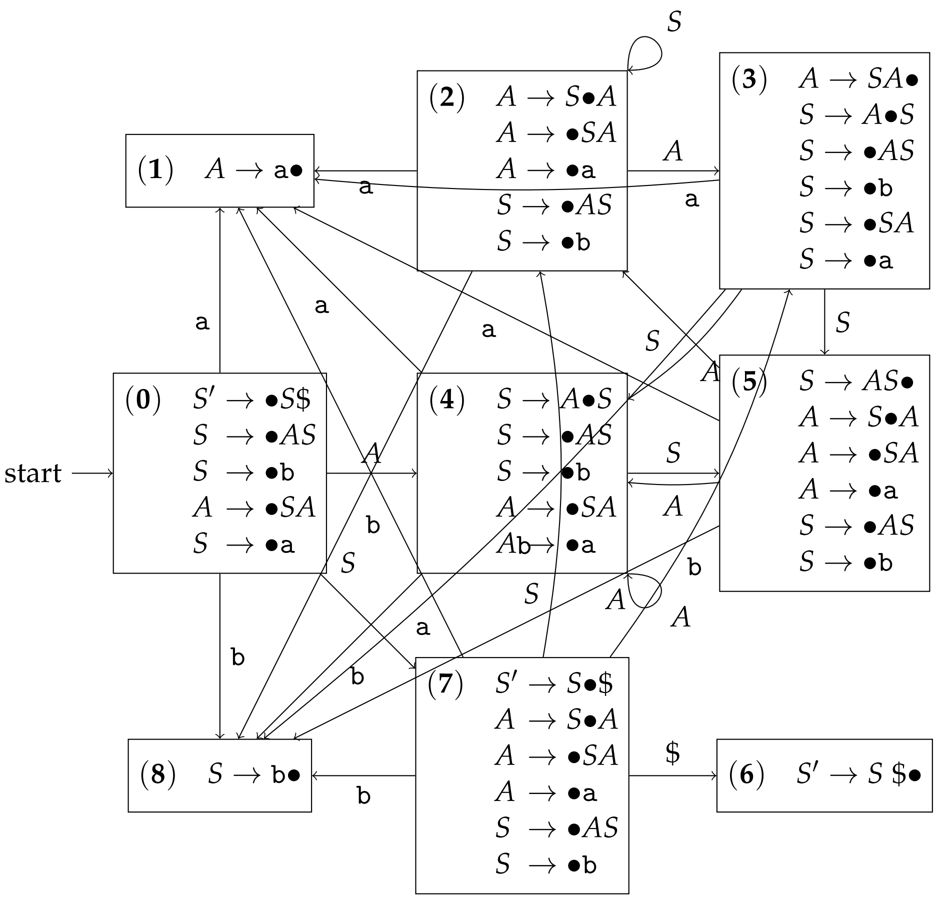
We want to use an LR parsing algorithm to parse sentences from this grammar. We start with extending the grammar with a new start-symbol S', and a production

$$S' \to S\ \$$$

where \$ is a terminal symbol denoting the end of input.

Construct the LR(0) automaton for the extended grammar. •

Solution 14. The LR(0) automaton corresponding to the full grammar looks as follows (each state is numbered before the production for future reference; the layout is not optimal, or, actually, terrible):



○