Department of Information and Computing Sciences
Utrecht University

# INFOB3TC – Exam 2

Johan Jeuring

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

## Preliminaries

- The exam consists of 12 pages (including this page). Please verify that you got all the pages.
- Fill out the answers **on the exam itself**.
- Write your **name** and **student number** here:

- The maximum score is stated at the top of each question. The total amount of points you can get is 100.
- Try to give simple and concise answers. Write readable text. Do not use pencils or pens with red ink. You may use Dutch or English.
- When writing grammar and language constructs, you may use any set, sequence, or language operations covered in the lecture notes.
- When writing Haskell code, you may use Prelude functions and functions from the following modules: *Data.Char*, *Data.List*, *Data.Maybe*, and *Control.Monad*. Also, you may use all the parser combinators from the `uu-tc` package. If you are in doubt whether a certain function is allowed, please ask.

*Good luck!*

## 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.

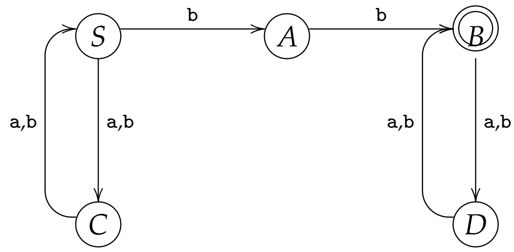
Answer these questions with *one of* a, b, c, or d.

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

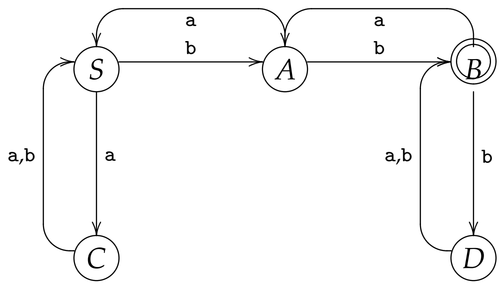
$$L \quad = \quad \{x \mid x \in \{\mathtt{a}, \mathtt{b}\}^*, \text{length } x \text{ is even}, \mathtt{bb} \text{ is a substring of } x\}$$

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

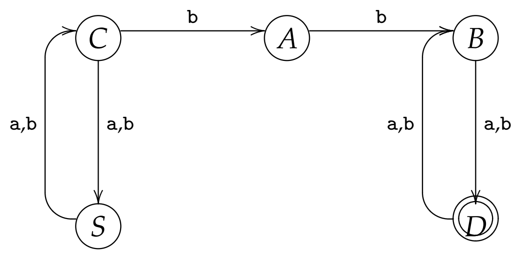
a)



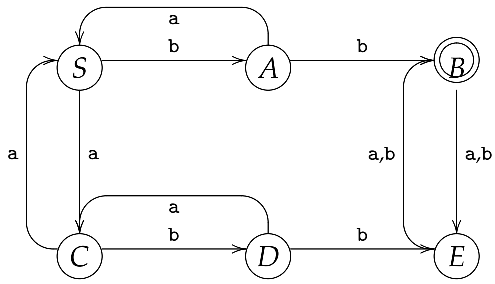

b)




c)

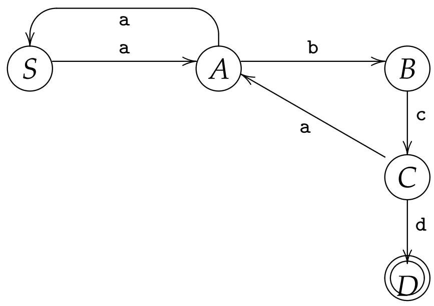



d)




•

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

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

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

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

**3** (5 points)**.** Consider the regular language

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

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

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

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

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

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

**4** (5 points). Consider the following two languages on the terminal symbols $\mathtt{x}$ and $\mathtt{y}$:

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

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.

•

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

•

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

$$\begin{aligned} L_1 &= \{0^n 1^m 0^n 1^m \mid n, m > 0\} \\ L_2 &= \{0^n 1^m 0^m 1^n \mid n, m > 0\} \end{aligned}$$

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.

•

| |
|---|
| |

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

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

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

a) $\{B\}$.

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

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

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

•

| |
|---|
| |

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

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

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

a) $\{B\}$.

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

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

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

•

| |
|---|
| |

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

I

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

II

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

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).

•

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

$$
\begin{array}{rcl}
S' & \rightarrow & S\$ \\
S & \rightarrow & X\mathtt{x} \mid Y\mathtt{y} \\
X & \rightarrow & \mathtt{x} \\
Y & \rightarrow & \mathtt{y}
\end{array}
$$

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.

•

| |
|---|
| |

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

$$
\begin{array}{rcl}
S & \rightarrow & E\$ \\
E & \rightarrow & A\mathtt{a}B \mid B \\
A & \rightarrow & \mathtt{b}B \mid \mathtt{c} \\
B & \rightarrow & A
\end{array}
$$

a) LR(1)

b) LALR(1)

c) SLR(1)

d) LR(0)

•

| |
|---|
| |

## 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. •

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

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

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.

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

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

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

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

$S' \rightarrow S\ \$$

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

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