The University of Nottingham

SCHOOL OF COMPUTER SCIENCE

A LEVEL 2 MODULE, SPRING SEMESTER 2014–2015

MACHINES AND THEIR LANGUAGES

ANSWERS

Time allowed TWO hours

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced.

Answer ALL THREE questions

No calculators are permitted in this examination.

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject-specific translation directories are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

Note: ANSWERS

Question 1

The following questions are multiple choice. There is at least one correct choice, but there may be several. To get all the marks you have to list all the correct answers and none of the wrong ones. 1 mistake yields 3 marks, 2 mistakes yield 1 mark, 3 or more yield zero.

Answer: *Note that the answer that should be provided is just a list of the correct alternative(s). The additional explanations below are just for your information.*

(a) Which of the following statements are correct?

(i) An alphabet is a finite set of symbols.

(ii) A word is a finite sequence of symbols over a given alphabet.

(iii) A language is an infinite set of words.

(iv) A finite language is never regular.

(v) A context-free language is never regular.

(5)

Answer: *Correct: i, ii*

Incorrect:

iii A language may be infinite, but does not have to be.

iv All finite languages are regular.

v All regular languages are also context-free.

(b) Which of the following statements are correct?

(i) For every language L, either ϵ is a symbol of the alphabet over which L is defined, or the empty word $\epsilon \notin L$.

(ii) $\epsilon \in \emptyset$

(iii) $\epsilon \in \emptyset^*$

(iv) The empty word ϵ belongs to all languages.

(v) For any language L, L^* is never the empty language.

(5)

Answer: *Correct: iii, v*

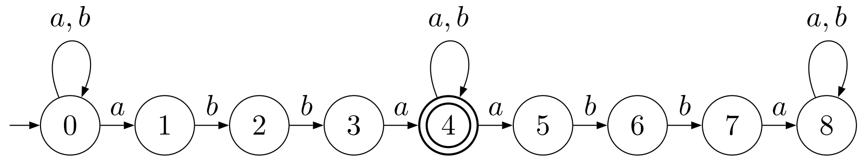
Incorrect:

i ϵ is never a symbol in an alphabet, and it is always possible to include the empty word in a language, regardless of the alphabet.

ii The empty language $\emptyset$, i.e. the empty set, does not contain any *words, not even the empty one.*

iv {a} is an example of a language that does not contain the empty word ϵ. An even simpler example is $\emptyset$.

(c) Consider the following finite automaton A over $\Sigma = \{a, b\}$:



Which of the following statements about A are correct?

(i) The automaton A is a Non-deterministic Finite Automaton (NFA).

(ii) $\epsilon \notin L(A)$

(iii) $ababbaba \in L(A)$

(iv) The language accepted by the automaton A is all words over Σ that contains the sequence *abba* at least once.

(v) The language accepted by the automaton A is all words over Σ that contains the sequence *abba* at most once.

(5)

Answer: *Correct: i, ii, iii, iv*

Incorrect:

v E.g. abbaabba $\in L(A)$.

(d) Consider the following regular expression:

$$(\mathbf{a} + \mathbf{ab} + \mathbf{abc})^*$$

Which of the following regular expressions denote the *same* language as the above regular expression?

(i) $(\epsilon + \mathbf{a} + \mathbf{ab} + \mathbf{abc})$

(ii) $\mathbf{a}(\epsilon + \mathbf{b} + \mathbf{bc})^*$

(iii) $\mathbf{a}^*(\epsilon + \mathbf{b} + \mathbf{bc})^*$

(iv) $(\mathbf{a}(\epsilon + \mathbf{b} + \mathbf{bc}))^*$

(v) $\mathbf{a}^* + (\mathbf{ab})^* + (\mathbf{abc})^*$

(5)

Answer: *Correct: iv*

Incorrect: i, ii, iii, v

(e) Consider the following Context-Free Grammar (CFG) G:

$$\begin{array}{rcl} S & \rightarrow & X \mid XY \\ X & \rightarrow & aXb \mid aYb \\ Y & \rightarrow & bYc \mid \epsilon \end{array}$$

where S, X, Y are nonterminal symbols, S is the start symbol, and a, b, c are terminal symbols.

Which of the following statements about the language $L(G)$ generated by G are correct?

(i) $\epsilon \in L(G)$

(ii) $aaabbbcc \in L(G)$

(iii) $aabbbbcc \in L(G)$

(iv) $\{a^i b^i b^j c^j \mid i, j \in \mathbb{N},\ i > 0\} = L(G)$

(v) The following CFG G' is equivalent to G above, i.e. $L(G') = L(G)$:

$$\begin{array}{rcl} S & \rightarrow & XY \\ X & \rightarrow & aXb \mid ab \\ Y & \rightarrow & bYc \mid \epsilon \end{array}$$

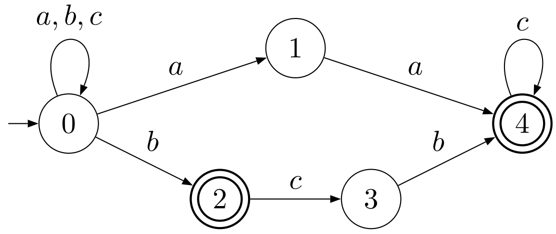
(5)

Answer: *Correct: iii*

Incorrect: i, ii, iv, v

Question 2

(a) Given the following NFA N over the alphabet $\Sigma = \{a, b, c\}$, construct a DFA $D(N)$ that accepts the same language as N by applying the *subset construction*:

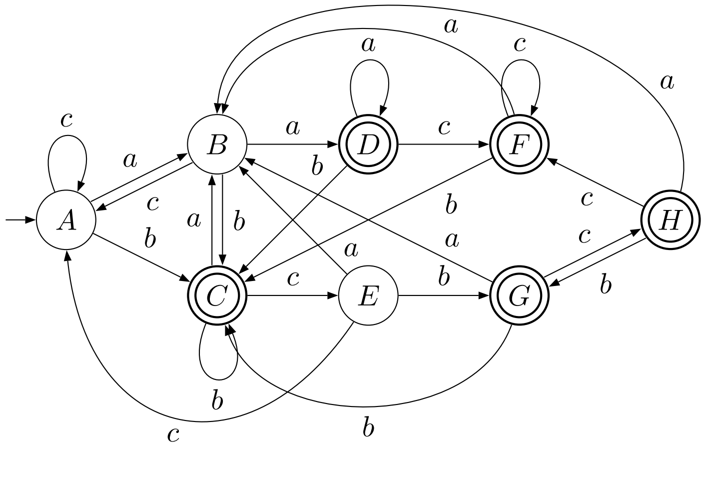


To save work, consider only the *reachable* part of $D(N)$. Clearly show your calculations in a state-transition *table*. Then draw the transition *diagram* for the resulting DFA $D(N)$. Do not forget to indicate the initial state and the final states both in the transition table and the final transition diagram. (13)

Answer:

	$\delta_{D(N)}$	a	b	c
$\rightarrow$	$\{0\} = A$	$\{0,1\} = B$	$\{0,2\} = C$	$\{0\} = A$
	$\{0,1\} = B$	$\{0,1,4\} = D$	$\{0,2\} = C$	$\{0\} = A$
$*$	$\{0,2\} = C$	$\{0,1\} = B$	$\{0,2\} = C$	$\{0,3\} = E$
$*$	$\{0,1,4\} = D$	$\{0,1,4\} = D$	$\{0,2\} = C$	$\{0,4\} = F$
	$\{0,3\} = E$	$\{0,1\} = B$	$\{0,2,4\} = G$	$\{0\} = A$
$*$	$\{0,4\} = F$	$\{0,1\} = B$	$\{0,2\} = C$	$\{0,4\} = F$
$*$	$\{0,2,4\} = G$	$\{0,1\} = B$	$\{0,2\} = C$	$\{0,3,4\} = H$
$*$	$\{0,3,4\} = H$	$\{0,1\} = B$	$\{0,2,4\} = G$	$\{0,4\} = F$

We can now draw the transition diagram for $D(N)$:



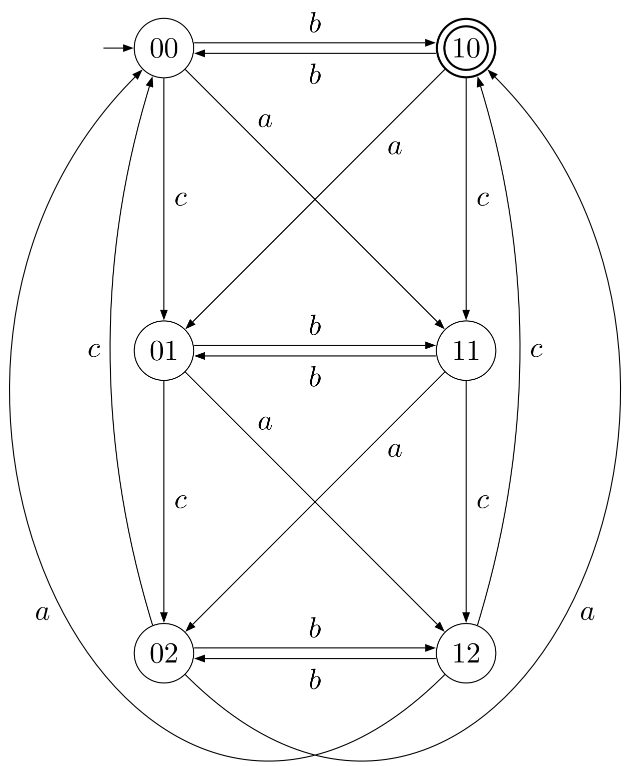
(b) For the alphabet $\Sigma = \{a, b, c\}$, construct a DFA C that recognises the words where the sum of the number of a's and the number of b's is odd, and the sum of the number of a's and c's is a multiple of three.

For example, $b \in L(C)$ (0 a's, 1 b, 0 c, $0 + 1 = 1$ is odd, $0 + 0 = 0$ is multiple of 3), $abbcc \in L(C)$ (1 a, 2 b's, 2 c's, $1+2 = 3$ is odd, $1+2 = 3$ is a multiple of three).

However, $\epsilon \notin L(C)$ (0 a's and 0 b's, $0+0 = 0$ is not odd), and $aaabbcc \notin L(C)$ (3 a's and 2 c's, $3 + 2 = 5$ is not a multiple of three).

Make sure to outline the idea behind your construction. (12)

Answer: *We need to keep track of the sum of the number of a's and b's counting modulo 2, and the sum of the number of a's and c's counting modulo 3. Let us label the states ij, where i is the sum of the number of a's and b's modulo 2 and j is the sum of the number of a's and c's modulo 3. We consequently get six states, with state 00 as initial state and state 10 the one accepting state. We then just have to consider what the resulting sums are when adding one more of each input symbol and add edges between the states accordingly. We thus obtain the following automaton:*



Question 3

(a) Consider the following Context-Free Grammar (CFG):

$$\begin{array}{rcl} S & \rightarrow & S \oplus A \mid A \\ A & \rightarrow & B \otimes A \mid B \\ B & \rightarrow & 0 \mid 1 \mid (S) \end{array}$$

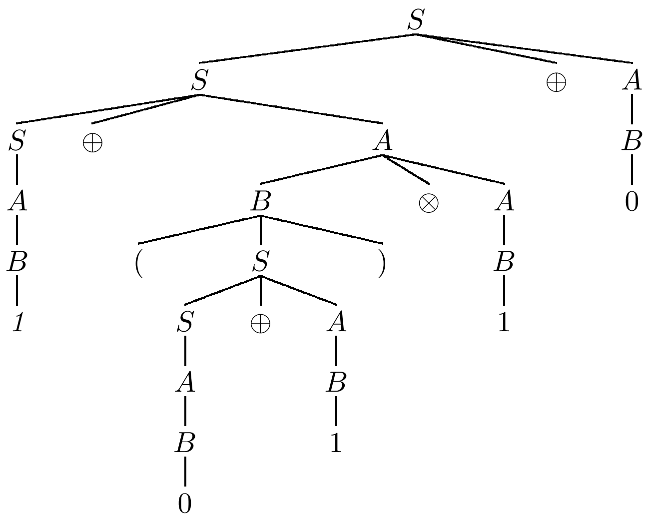
S, A, and B are nonterminals; 0, 1, $\oplus$, $\otimes$, (, and) are terminals; S is the start symbol.

Draw the derivation tree according to this grammar for the word

$$1 \oplus (0 \oplus 1) \otimes 1 \oplus 0$$

(6)

Answer: *Derivation tree for* $1 \oplus (0 \oplus 1) \otimes 1 \oplus 0$*:*



(b) Is the following CFG ambiguous? If yes, show this. If no, explain why.

$$\begin{array}{rcl} A & \rightarrow & aBbA \mid aBbAcA \mid d \\ B & \rightarrow & e \end{array}$$

A and B are nonterminals, A is the start symbol, a, b, c, d, and e are terminals. (9)

Answer: *Yes, the grammar is ambiguous. For example, two different* *leftmost derivations for the word aebaebdcd:*

$$\begin{aligned} A &\underset{\text{lm}}{\Rightarrow} aBbA \\ &\underset{\text{lm}}{\Rightarrow} aebA \\ &\underset{\text{lm}}{\Rightarrow} aebaBbAcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebAcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebdcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebdcd \end{aligned}$$

and

$$\begin{aligned} A &\underset{\text{lm}}{\Rightarrow} aBbAcA \\ &\underset{\text{lm}}{\Rightarrow} aebAcA \\ &\underset{\text{lm}}{\Rightarrow} aebaBbAcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebAcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebdcA \\ &\underset{\text{lm}}{\Rightarrow} aebaebdcd \end{aligned}$$

(c) Consider the following Context-Free Grammar (CFG):

$$\begin{aligned} S &\rightarrow ABC \mid BC \\ A &\rightarrow aA \mid a \\ B &\rightarrow b \mid C \\ C &\rightarrow cc \mid dd \mid \epsilon \end{aligned}$$

S, A, B, and C are nonterminals, a, b, c, and d are terminals, and S is the start symbol.

(i) What is the set N_ϵ of *nullable* nonterminals for this grammar? Provide a brief justification. (3)

Answer: *$N_\epsilon = \{S, B, C\}$. C is nullable because $C \rightarrow \epsilon$ is a production. B is nullable because $B \rightarrow C$ is a production and C is nullable. S is nullable because $S \rightarrow BC$ is a production and both B and C are nullable. A is not nullable since the RHS of all productions for A start with a terminal (a).*

(ii) Systematically compute the *first sets* for all nonterminals, i.e. first(S), first(A), first(B), and first(C), by setting up and solving the equations according to the definitions of first sets for nonterminals and strings of grammar symbols. Show your calculations. (7)

Answer:

$$\begin{aligned}
\text{first}(A) &= \text{first}(aA) \cup \text{first}(a) \\
&= \{a\} \cup \{a\} \\
&= \{a\}
\end{aligned}$$

$$\begin{aligned}
\text{first}(C) &= \text{first}(cc) \cup \text{first}(dd) \cup \text{first}(\epsilon) \\
&= \{c\} \cup \{d\} \cup \emptyset \\
&= \{c, d\}
\end{aligned}$$

$$\begin{aligned}
\text{first}(B) &= \text{first}(b) \cup \text{first}(C) \\
&= \{b\} \cup \{c, d\} \\
&= \{b, c, d\}
\end{aligned}$$

$$\begin{aligned}
\text{first}(S) &= \text{first}(ABC) \cup \text{first}(BC) \\
&= \text{first}(A) \cup (\text{first}(B) \cup \text{first}(C)) \\
&= \{a\} \cup \{b, c, d\} \cup \{c, d\} \\
&= \{a, b, c, d\}
\end{aligned}$$