Lahore University of Management Sciences
Department of Computer Science

Roll #: SOLUTION
Total Pages: 09 (including this page)
Quarter: Winter
Academic Year: 2007-2008
Date: January 21, 2008
Time Allowed: 100 minutes
Total Marks: 75

Course Title: Automata Theory
Course Code: CS 311 / MA 352
Instructor: Shahab Baqai
Exam: Midterm

The instructions below must be followed strictly. Failure to do so can result in serious grade loss.

DO NOT OPEN THIS EXAM UNTIL TOLD TO DO SO.

⇒ Read all questions very carefully before answering.
⇒ Check the number of papers in the question sheet and make sure the paper is complete.
⇒ Keep your eyes on your own paper.
⇒ You may not
    • Talk to anyone once the exam begins.
    • Leave the examination room and then return.
    • Use any communication device

Specific instructions:

<table>
<thead>
<tr>
<th>Open book/notes, help sheet:</th>
<th>Closed Book &amp; Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calculator usage:</td>
<td>None</td>
</tr>
<tr>
<td>Write in pen/pencil:</td>
<td>Optional</td>
</tr>
<tr>
<td>Any other instruction(s):</td>
<td>no communication devices, e.g. cell phones etc.</td>
</tr>
</tbody>
</table>

Score

<table>
<thead>
<tr>
<th></th>
<th>Q #1</th>
<th>Q #2</th>
<th>Q #3</th>
<th>Q #4</th>
<th>Q #5</th>
<th>Q #6</th>
<th>Q #7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>10</td>
<td>15</td>
<td>05</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td>Awarded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Q # 1 (10 points)

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. Which of the following statements are correct?
   a. An alphabet is a finite set of symbols.
   b. A string is a possibly infinite sequence of symbols over a given alphabet.
   c. A language is the set of all possible strings over a given alphabet.
   d. A regular language is always finite.
   e. A finite language is always regular.

2. Which of the following statements are correct?
   a. The empty string $\lambda$ belongs to all languages.
   b. The empty string $\lambda$ is the only word in the empty language $\phi$.
   c. $\lambda \in \{a^*\}$
   d. $\lambda \in \phi^*$
   e. If $L$ is a language containing at least one non-empty word, then $L^*$ is an infinite language.

3. Consider the finite automaton $A$ over $\Sigma = \{a, b\}$
   Which of the following statements about $A$ are correct?
   a. The automaton $A$ is a Deterministic Finite Automaton (DFA).
   b. $\lambda \in L(A)$
   c. $abba \in L(A)$
   d. All words accepted by $A$ contain equally many $a$'s and $b$'s.
   e. The automaton $A$ accepts all non-empty words over $\Sigma$ that contain equally many $a$'s and $b$'s.

4. Consider the set $W = \{\lambda, ab, cab, abab\}$. Which of the following regular expressions denote a language that contains all strings in $W$? (But not necessarily only the strings in $W$ – It is OK if the language denoted by the regular expression contains more words.)
   a. $(\lambda + ab + c) (\lambda + ab)$
   b. $(\lambda + ab + c) (\phi + ab)$
   c. $(\lambda + ab + c)^*$
   d. $(ab + c)^*$
   e. $(ab\phi + c)^*$
   f. $(ab)^* + c^*$
5. Consider the following Context-Free Grammar (CFG) G:

\[
S \rightarrow X \mid Y C \\
X \rightarrow aXc \mid B \\
Y \rightarrow aYb \mid \lambda \\
B \rightarrow bB \mid \lambda \\
C \rightarrow cC \mid \lambda.
\]

S, X, Y, B, C are variables, 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?

a. \( aabcc \in L(G) \)

b. \( \{ a^i b^j c^k \mid i, j, k \in \mathbb{N} \} \subseteq L(G) \)

c. \( \{ a^i b^j c^k \mid i, k \in \mathbb{N} \} \subseteq L(G) \)

d. \( \{ a^i b^j c^i \mid i \in \mathbb{N} \} \subseteq L(G) \)

e. \( L(G) = \{ a^i b^j c^i \mid i \in \mathbb{N} \} \)

\[ L(G) = \{ a^2 b^2 c^2 \} \cup \{ a^i b^i c^i \} \]
Q #2 (15 points)

True/False with justification:
1. Complement of a CFL may also be CFL.
   **True.** All regular languages are CFL and their complement is also regular.

2. Subset of a regular language is always regular.
   **False.**
   \[
   L_1 = a^* b^* \\
   L_2 = a^n b^n \\
   L_2 \subseteq L_1
   \]

3. While converting an NFA into a DFA, each state of the DFA represents an equivalence class of states of NFA.
   **False.** Each state of DFA represents a possible subset of states of NFA.

4. Pumping lemma is a property obeyed by all non-regular languages and therefore we use it to prove a language to be non-regular.
   **False.** Pumping lemma is a property of regular languages and used to prove a language non-regular by contradicts.

5. In $R_{ij}^k$ notation of constructing a regular expression, $i$ is the source state, $j$ is the target state and $k$ is maximum number of hops to be made from $i$ to $j$.
   **False.** $k$ is the maximum possible label of an intermediate state between $i$ and $j$.

6. Given a minimized DFA, it is possible to construct an NFA for the same language with less number of states.
   **True.**

   \[
   \text{DFA: } \quad \begin{array}{c}
   \text{NFA: }
   \end{array}
   \]

7. Given an NFA having $n$ states, its equivalent DFA will have at least $n$ states.
   **False.**
Q #3 (05 points)
Let $\Sigma = \{a, b\}$. Let $L$ be the set of strings that do not have the word $aab$ as a (contiguous) substring
\[ L = \{ w \mid w \neq w_1aabw_2 \text{ for any } w_1, w_2 \in \Sigma^* \} \]

For example, $baa \in L$ and $abab \in L$, but $aab \notin L$, $babaabab \notin L$, and $ababaab \notin L$. Note that $L$ contains all strings shorter than 3 characters.

Construct a DFA that accepts $L$ and give a state diagram showing all states in the DFA. Your DFA should not contain more than 10 states.

Explain briefly how your DFA works. You do not need to prove formally that your DFA is correct.
Q # 4 (10 points)

Let $\Sigma = \{0, 1\}$. Let $L$ be the language containing all strings over $\Sigma$ that have length at least three and have a 0 in the third position from the end.

That is, $L = \{w_1c_1c_2 \mid w_1 \in \Sigma^*; \ c_1, c_2 \in \Sigma\}$

For example, $L$ contains 001000 and 11011, but $L$ does not contain 00 or 0100.

Give the state diagram of an NFA that recognizes $L$ and has no more than 6 states. Full credit requires a machine that isn’t overly complex and that exploits the extra features of NFAs. Don’t solve this with a DFA.
Q #5 (15 points)

Let $J = \{ww^R \mid w \in \{a,b\}^*, |w| \geq 1\}$, where $w^R$ is the reversal of the string $w$. Notice that $J$ does not contain the empty string.

Define the language $L$ by $L = \{a^n \#x_1 \#x_2 \ldots \#x_n \mid n \geq 1, x_i \in J \text{ for all } i\}$.

That is, any string in $L$ starts with an initial $a^n$, which is a counter telling you how many items from $J$ to expect in the rest of the string. For example, $a\#abba$ and $aa\#babbab\#bb$ are both in $L$. But none of the following are in $L$: $\epsilon$, $a\#abba\#bbbbb$, $a\#abab$, $a\#aba$.

(a) Give a context-free grammar whose language is $J$.

The grammar has the start symbol $T$.

$$T \rightarrow aTa \mid bTb \mid aa \mid bb$$

Common mistakes: $x \not\in J$

$$a \not\in J$$

$$b \not\in J$$

(b) Give a context-free grammar whose language is $L$.

The grammar has the start symbol $S$.

$$S \rightarrow aS \#T \mid a \#T$$

$$T \rightarrow aTa \mid bTb \mid aa \mid bb \quad (\text{from (a)})$$

It is important that a single recursive step generates an "$a$" and a "$T$" at the same time, so the number of $a$'s will match the number of $T$'s.
Q # 6 (10 points)

Let \( L = \{ c^a a^i b^j c^a \mid n \geq 0, j > i \geq 0 \} \). Notice that \( n \) and \( i \) can be zero, but \( j \) must be greater than or equal to 1.

Give the state diagram for a PDA whose language is \( L \). Include brief comments explaining the design of your PDA, to help us understand how it works.

The last transition checks when the start symbol is at the top so as to move to the accept state, ensuring that the numbers of 'c' match.

States 9, and 92 push a single 'c' onto the stack for each pair of input 'c's.

State 93 pushes meaning 'c' onto the stack

State 94 compares meaning b's to the a's on the stack

State 95 reads any extra b's & the transition from state 95 ensures that at least one extra b has been read.
7 (10 points)
Determine or disprove that following languages are context-free:
we {a, b}*: Number of a’s is divisible by 3

Hence, L₁ is regular and all regular languages are CFL.