Instructions: There are six questions and each question is worth 10 marks. Midterms will be marked out of 50 based on your top five answers from (1)-(6). Time limit is 90 minutes.

(1) (10 marks) In this question you give a regular expression and an $\epsilon$-finite automata.

(a) (5 marks) Give a RE for the binary strings with an even number of 1s.

(b) (5 marks) Give an $\epsilon$-FA that accepts the following language

$L = \{ w \in \{a, b, c\}^* | w \text{ contains } ab \text{ as a substring and also } ca \text{ as a substring} \}$.

(a) $0^* (10^* 10^*)^*$ another solution is $(0+10^*)^*$

(b) $\begin{aligned}
\text{(substring } ab \text{ then } ca) \\
\text{(substring } cab) \\
\text{(substring } ca \text{ then } ab)
\end{aligned}$

[This can also be done easily with fewer states and no $\epsilon$-transitions]
(2) (10 marks) This question focuses on the pumping lemmas.
(a) (5 marks) State the pumping lemma for context-free languages.
(b) (5 marks) Prove that the following language is not regular by using the pumping
lemma for regular languages
\[ \mathcal{L} = \{ w \in \{0, 1, 2, 3\}^* | w \text{ contains more } 1s \text{ than } 2s \} \].

(a) For every context-free language \( \mathcal{L} \), there exists a constant \( p \) such that
for all \( s \in \mathcal{L} \) with \( |s| \geq p \), there exist \( uvwxy \) such that
(i) \( s = uvwxy \),
(ii) \( |vwx| \leq p \),
(iii) \( |vx| > 0 \), and
(iv) \( uv^iwx^iy \in \mathcal{L} \) for all \( i \geq 0 \).

(b) Consider \( w = 1^{n+1}2^n \). Observe \( w \in \mathcal{L} \) and \( |w| > n \).

Suppose \( x, y, z \) are chosen such that (i) \( w = xyz \), (ii) \( |xy| \leq n \),
and (iii) \( |y| > 0 \). Under this assumption it must be that
\( x = 1^j \), \( y = 1^k \), and \( z = 1^{n+1-j-k}2^n \) where \( j \geq 0 \), \( k > 0 \), and \( j + k \leq n + 1 \).

Observe that when \( i = 0 \), \( xy^iz = xy = 1^{n+1-k}2^n \notin \mathcal{L} \) since \( k > 0 \)
implies that \( n+1-k \leq n \), and so the string does not have more \( 1s \)
than \( 2s \). Therefore, \( \mathcal{L} \) does not satisfy the (regular) pumping lemma and hence is not regular.
(3) (10 marks) For this question, you will recall the definition of the two languages accepted by a PDA $P = (Q, \Sigma, \Gamma, \delta, q_0, Z_0, F)$.

(a) (2 marks) What are the components of an instantaneous description (ID) $(, , )$?

(b) (2 marks) Define $\vdash$ for two IDs. In other words, when does $(, , ) \vdash (, , )$ hold?

(c) (2 marks) Define $\vdash^*$. In other words, when does $A \vdash^* B$ hold for IDs $A$ and $B$?

(d) (2 marks) Define $L(P)$, the language accepted by final state.

(e) (2 marks) Define $N(P)$, the language accepted by empty stack.

(a) $(q, w, \alpha)$ where $q \in Q$ is a state, $w \in \Sigma^*$ is the input string remaining to be read, and $\alpha \in \Gamma^*$ is the current content of the stack.

(b) If $(q, \beta) \in \delta(p, a, X)$ for $p, q \in Q$, $\beta \in \Gamma^*$, $a \in \Sigma U \{\epsilon\}$, $X \in \Gamma$ then $(p, aw, X\alpha) \vdash (q, w, \beta\alpha)$ for all $w \in \Sigma^*$ and $\alpha \in \Gamma^*$.

(c) If $A \vdash^* C$ and $C \vdash B$ then $A \vdash^* B$ for all IDs $A, B, C$.

(d) $L(P) = \{ w \in \Sigma^* | (q_0, w, Z_0) \vdash^* (q, \epsilon, \alpha) \text{ for some } q \in F \text{ and } \alpha \in \Gamma^* \}$

(e) $N(P) = \{ w \in \Sigma^* | (q_0, w, Z_0) \vdash^* (q, \epsilon, \epsilon) \text{ for some } q \in F \}$
(4) (10 marks) Provide a PDA that accepts the following language:

\[ L = \{ w \in \{0, 1, 2\}^* \mid w = 0^i1^j2^k \text{ and } k \geq i + j \}. \]

Specify whether your PDA accepts by final state or empty stack. You do not need to prove that your PDA is correct.

\[ P = (Q, \Sigma, \Gamma, \delta, q_0, F) \]

\[ Q = \{ q_0, q_1, q_2 \}, \Sigma = \{0, 1, 2\}, \Gamma = \{(0, 1, 2)\}, \delta, q_0, F \]
(5) (10 marks) This question is on ambiguous context-free grammars. The first three parts focus on the CFG $G_b = (\{S\}, \{0, 1\}, P, S)$ with productions $P$ given below

$$S \rightarrow S1S0S$$
$$S \rightarrow e.$$ 

(a) (2 marks) What is $L(G_b)$? You do not need to prove your result.

(b) (2 marks) Prove that $G_b$ is ambiguous.

(c) (3 marks) Show that $L(G_b)$ is not ambiguous by providing an unambiguous CFG $G$ such that $L = L(G)$. You do not need to prove that $G$ is unambiguous.

(d) (3 marks) A CFL $L$ is inherently unambiguous if the following property holds: If $G$ is a CFG such that $L = L(G)$, then $G$ is unambiguous. Provide an inherently unambiguous CFL that is non-empty ($L \neq \emptyset$) or prove that no such CFL exists.

(a) $L(G_b) = \{w \in \{0, 1\}^* \mid \text{every prefix of } w \text{ contains at least as many } 1\text{s as } 0\text{s, and } w \text{ contains an equal number of } 1\text{s and } 0\text{s} \}$

In other words, balanced parentheses with 1 for '(' and 0 for ')'.

(b) There are two distinct parse trees for 1010: 1 so $L$ is ambiguous.

(c) $S \rightarrow 1S0S \mid e$

(d) If $L$ is a CFL, then there exists a CFG $G = (V, \Sigma, P, S)$ such that $L(G) = L$. Now consider the CFG $G' = (V \cup \{A', B'\}, \Sigma, P', S')$ where $P'$ contains all productions in $P$ and the following additional rules:

$$S' \rightarrow A' \mid B'$$
$$A' \rightarrow S$$
$$B' \rightarrow S$$

Observe $L(G') = L(G) = L$ and $G'$ is ambiguous since $S \Rightarrow A' \Rightarrow S$ or $S \Rightarrow B' \Rightarrow S$. A non-empty inherently unambiguous
(6) (10 marks) This question asks you to transform a non-deterministic finite automaton into an equivalent finite automaton.

(a) (8 marks) Transform the NFA $M$ given below into an equivalent FA using the accessible subset construction.

(b) (2 marks) Is there a single transition in $M$ that can be removed without changing the accessible subsets of states? Provide the transition, or state that none exists.

(a) | subsets | 0 | 1 |
---|---|---|---|
(0) $\{a, b\}$ | $\{a\}$ | $\{a, b\}$
(01) $\{a, c\}$ | $\{a, b, d\}$ | $\{a, b\}$ **
(010) $\{a, b, d\}$ | $\{a, b, d\}$ | $\{a, c\}$

(b) The transition $\overset{1}{\text{(b)}} \overset{0}{\text{(c)}}$ can be removed without changing the accessible subsets. In particular, the only entry in the table from (a) that changes is ** to $\{a, b\}$. 

...