Attach a fully filled-in cover sheet to the front of your printed homework. Your name should not appear anywhere; the cover sheet and each individual page of the homework should include your Banner ID only.

While collaboration is encouraged in this class, please remember not to take away notes from any labs or collaboration sessions.

Please monitor Piazza, as we will post clarifications of questions there. Hand in your solutions by 12:55 to the CSCI 1010 bin on the second floor of the CIT. Late homeworks are not accepted.

**Problem 1**

In Will Owens’s continuing investigation of the Illuminatree, he begins to suspect that they have influential members all throughout the Pinary Tree Company. In order to communicate his work with his supervisors Douglas and Connie Ferr, he and the Ferrs establish their own private language. Will is tasked with building an interpreter to determine the veracity or fradulence of a purported communiqué.

All valid messages are in the language $L$ of the regular expression:

$$(10)^*(011 \cup 1)^*$$

Any message outside this language (that is, a message in the language $L^C$) is assumed to be the work of covert Illuminatree agents or ne’er-do-well bootloggers. Help Will build the interpreter by providing a DFSM whose language is $L$.

**Problem 2**

Consider strings over the alphabet $\{0, 1\}$. Provide a regular expression for the complement of the language $1^*$. Show your work.
Problem 3

In the midst of the Pinary Tree Company’s troubles with the Illuminatree, the Ferrs have an even more momentous concern: the Game of Cones. The Game of Cones is an annual contest between all of the conifer-centric organizations in Yew Nork. The producer of the highest quality conifer products is awarded a prized trophy, the Iron Cone. In past years, competition has grown so fierce that opposing factions of conifer enthusiasts have taken to poisoning one another’s arboreta with those disgusting deciduous trees. This year, the mayor of Yew Nork steps in to mandate that everyone use the same community arboretum for the contest.

The Ferrs are assigned to the same area as U. K. Liptus. The Ferrs get the odd spaces and U. K. Liptus gets the even spaces. To further quell competition, the mayor requires that the Ferrs and U. K. Liptus grow exactly the same number of trees. Begrudgingly, the three of them work together to plan their planting area. They need to leave some empty space for the trees’ roots to grow, so they represent the area as a binary string, with 1’s representing conifer trees and 0’s representing empty spaces. The planting arrangement is only acceptable if the Ferrs and U. K. Liptus each grow the same number of trees, where the Ferrs’ are in the odd indices and U. K. Liptus’s are in the even indices.

More formally, the language of acceptable arboreta, \( L \), is defined as:

\[
L = \{ w \mid w \text{ contains the same number of 1’s in even and odd indices} \}
\]

Use the pumping lemma to show that the language of acceptable arboreta is not regular.
The following questions are lab problems.

Lab Problem 1

Recall that a **pumping length** for a language $A$ is a positive integer $p$ such that all strings $s \in A$ with $|s| \geq p$ can be written in the form $xyz$, where:

(i) $|xy| \leq p$,
(ii) $|y| \geq 1$,
(iii) and $xy^iz \in A$ for all $i \geq 0$.

Also, recall from class that if $A$ is finite with its longest string of length $\ell$, then $p = \ell + 1$ is a valid pumping length for $A$ because there are no strings $s \in A$ with $|s| \geq \ell + 1$. This makes it vacuously true that all such strings satisfy the three conditions above.

The **pumping lemma** states that every regular language has a pumping length. The **minimum pumping length** of a language $A$, $p_{\text{min}}$, is the smallest pumping length for $A$. Note that this implies that every integer $p \geq p_{\text{min}}$ is also a valid pumping length for $A$.

For example, if $A = ab^*$, the minimum pumping length is two. To justify this, note that the string $s = a$ is in $A$ but cannot be pumped at all. Writing it as $a = xyz$, we must have $x = \varepsilon$, $y = a$, and $z = \varepsilon$, but then $xz$ is not in $A$. This shows that 1 is not a pumping length. However, 2 is a pumping length because for any string $|s| \geq 2$, we can take $x = a$, $y = b$, and $z$ to be everything else, so that we have $|xy| \leq 2$, $|y| \geq 1$, and $xy^iz \in A$ for all $i \geq 0$.

For each of the following languages, give the minimum pumping length $p_{\text{min}}$ and prove your answer.

a. $bb^*$

b. $ba(bb^*a)^*a$

c. $L_c = \{ w | w \in \{a, b\}^* \text{ and } w \text{ ends in } aa \}$

d. $L_d = \{aabb, bbba, abab, a\}$

e. $L_e = \{ w | w \in \{a, b\}^* \text{ and } w \text{ does not end in } aa \}$
Lab Problem 2

Not to be outmanoeuvred by the Ferrs and Will Owens—and to boost their own members’ chances in the Game of Cones—the Illuminatree decides to invent their own secret language. Now that everyone in the Game of Cones needs to share community arboreta, it is especially imperative to safeguard sensitive stratagems. The Illuminatree’s language is so secret, in fact, that no FSM can recognize it.

Their language, \( F \), is defined over the alphabet \( \Sigma = \{a, b, c\} \) as:

\[
F = \{a^i v \mid i \geq 0, v \text{ is a sequence of } b\text{'s and } c\text{'s, and if } i = 1 \text{ then } v \text{ is of the form } xx: \text{ it consists of two identical strings}\}
\]

a. Prove that \( F \) is not regular.

b. Give a pumping length \( p \) and demonstrate that for all strings \( w \in F \) such that \( |w| \geq p \), we can write \( w = xyz \) with \( |xy| \leq p \) and \( |y| \geq 1 \) such that \( xy^iz \in F \) for all \( i \geq 0 \). In other words, the pumping lemma is not helpful for proving that \( F \) is not regular.

c. Explain why parts (a) and (b) do not contradict the pumping lemma.

Lab Problem 3

In this problem, you will prove that regular expressions, DFMSs, and NFSMs all decide the same set of languages. To do this, we introduce the concept of a generalized NFSM (GNFSM). A GNFSM is simply an NFSM whose transition inputs are regular expressions instead of individual symbols.

A GNFSM \( G \) is a 5-tuple of the form \( G = (\Sigma, Q, \delta, s, f) \). The input alphabet is \( \Sigma \), the set of states is \( Q \), the start state is \( s \in Q \), the accepting state is \( f \in Q \), and \( \delta : (Q \setminus \{f\}) \times R \to (Q \setminus \{s\}) \) is the state transition function defined with \( R \) as the set of all regular expressions over \( \Sigma \).

The definition of \( \delta \) for GNFSMs implies that there are no outgoing edges from the accepting state and no incoming edges to the start state. Moreover, we require that the start state have transitions to every other state, and that the accepting state have transitions from every other state. The accepting state is not the same as the start state. Finally, among all non-start and non-accepting states, there is one transition from every state to every other state, including itself.
Here is an example of a GNFSM:\footnote{Illuminatree confirmed.}:

\begin{center}
\includegraphics[width=\textwidth]{gnfsm_diagram.png}
\end{center}

In GNFSM diagrams, undrawn transitions are assumed to be labeled with the empty set. You may assume without proof that a GNFSM can be converted to a regular expression and to an NFSM.

a. Show that a DFSM can be converted to a GNFSM.

b. Show that an NFSM can be converted to a GNFSM.

c. Show that a regular expression can be converted to a GNFSM.

d. Conclude that regular expressions, DFSMs, and NFSMs recognize the same set of languages.