Comp 330: Mid-term examination

Instructor: Jérôme Waldispühl

October 29, 2023.

- Answer the questions on Crowdmark.
- You have 1 hour 10 minutes to complete this exam and 10 extra minutes to upload your answers and account for any technical issue. You must initiate your submission within 1 hour and 10 minutes from the start. If you meet any difficulty while uploading your solution 5 minutes before the end of the submission timeframe (1h + 20min), email the instructor at `jerome.waldispuhl@mcgill.ca` with your solution.
- You can update your submission on Crowdmark during the timeframe of the exam.
- The clarity and presentation of your answers is part of the grading. Answers poorly presented may not be graded. This includes the clarity of the writing or the quality of the image you uploaded. It is your responsibility to ensure that the image has the good resolution and contrast.
- Keep the size of any file you upload on Crowdmark as small as possible to avoid technical issues.
- Unless specified, all answers must be explained.
- Partial answers will receive credits.
- The conciseness of your answer is part of the grading. An answer that is unnecessarily long or poorly structured will be penalized.
- This is an open book examination.
- It is strictly forbidden to use any external help, including online tutoring systems, or to provide aid to someone else. You are not allowed to communicate to anyone during the exam.
- It is strictly forbidden to share or disseminate this exam or any information related to this exam.
- This exam contains a mandatory academic integrity statement that you should agree with and sign. *We will not grade the exam otherwise*.
- This exam contains 8 pages.

Full name: ______________________________

Student ID: ______________________________

Statement of Academic Integrity

In submitting this exam, I confirm that my conduct during this exam adheres to the Code of conduct and academic integrity (`https://www.mcgill.ca/students/srr/academicrights`). I confirm that I did NOT act in such a way that would constitute cheating, misrepresentation, or unfairness, including but not limited to, using unauthorized aids and assistance, personating another person, and committing plagiarism. I will not share or disseminate this exam on any platform or through personal communication.

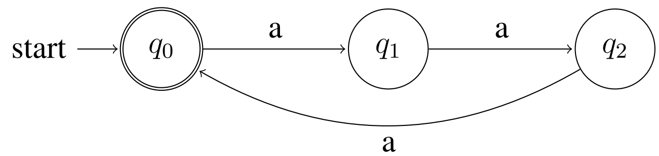
Write your name and date to sign this statement. *We will not grade your exam if you do not agree with and sign this statement.*

Regular Languages

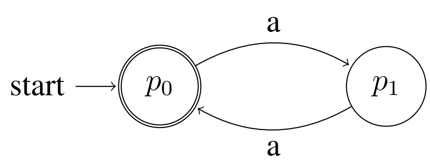
1. Let M be the language defined by the regular expression $(aaa)^* - (aa)^*$.

   (a) (10 points) Build a deterministic finite automaton (DFA) accepting M (Hint: Start by building a DFA accepting the regular language $K = (aaa)^*$ and another DFA accepting the regular language $K = (aa)^*$. Then, note that $M = K - L = K \cap \bar{L}$).

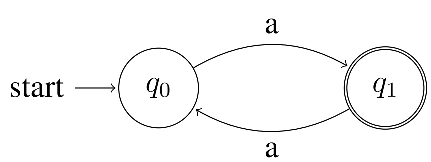
   Solution: The DFA for $K = (aaa)^*$:

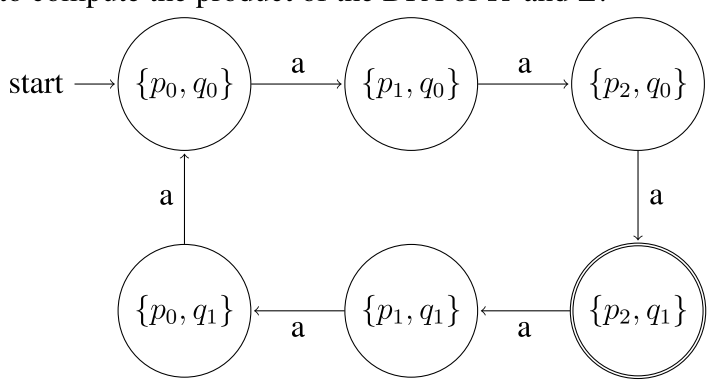
   Then, the DFA for $L = (aa)^*$:

   And the DFA for $\bar{L}$:

   We then just need to compute the product of the DFA of K and $\bar{L}$:

   (b) (5 points) Is the language M empty? Is it finite? Justify your answers.

   Solution: M is not empty because the automaton has a path from the start state to an accept state. The language also has an infinite number of words because the automaton has a cycle accessible from the initial state.

   (c) (5 points) Using your automaton, find a (simple) regular expression for M.

   Solution: $a^3(a^6)^*$

2. (10 points) Using the pumping lemma for regular languages, show that the language L of words that are **not** palindromes (i.e., a palindrome is a string that reads the same forward and backward) is not a regular language (Hint: use the pumping lemma to show that the complement of L is not regular).

Solution: First, we show that the language P of palindrome is not regular. Let k be the number of the pumping lemma and let $x = a^k b a^k$ be a work of P such that $|x| \geq k$. Therefore, $\exists u, v, w$ such that $x = uvw$ with $|uv| \leq k$ and $|v| \geq 1$. The pumping lemma states that $\forall i \geq 0, uv^i w \in P$. Because $|uv| \leq k$, then it occurs before the only b and must be entirely made of a's. Since $|v| \geq 1$, then uw has at least one less a before the only b than after. The word cannot be a palindrome. Contradiction. Thus P is not regular.
Finally, we conclude that L cannot be regular because otherwise $\bar{L} = P$ would be regular by the closure property of the complement of a language.

Context-Free Languages

3. Let T be the language such $T = \{a^i b^j \mid i \neq j\}$

   (a) (10 points) Show that T is a context-free language (CFL) by proposing a context-free grammar (CFG) that generates its (Hint: Start with CFGs generating the CFL's $U = \{a^i b^j \mid i < j\}$ and $V = \{a^i b^j \mid i > j\}$).

   Solution: The language $U = \{a^i b^j \mid i < j\}$ can be generated with the following production rules.

   $$A \rightarrow aAb \mid Ab \mid b$$

   The language $V = \{a^i b^j \mid i > j\}$ can be generated with the following production rules.

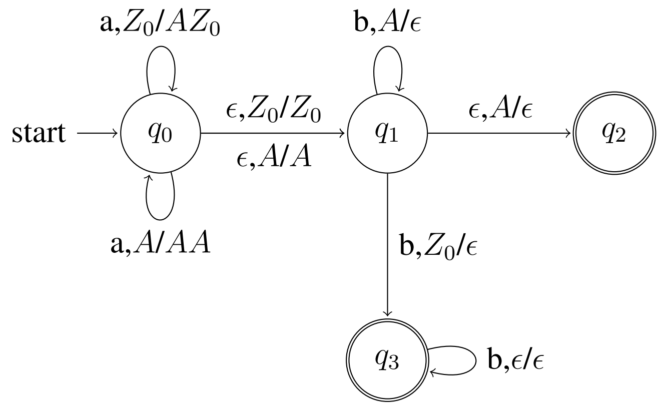
   $$B \rightarrow aBb \mid aB \mid a$$

   Since $T = U \cup V$, we can build a grammar generating T using the production rules from above and the following ones (where S is the new start symbol):

   $$S \rightarrow A \mid B$$

   (b) (10 points) Find a PDA accepting T.

   Solution: We build the following PDA such that the state q_2 accepts the words that have more a's than b's, while q_3 accepts the word that have more b's than a's.



(c) (10 points) What is the language $a^*b^* - T$?

Solution:

$$a^*b^* = \{a^n b^m \mid n, m \in \mathbb{N}\}$$

$$a^*b^* - T = \{a^n b^m \mid n, m \in \mathbb{N} \wedge \neg(n \neq m)\} = \{a^i b^i \mid i \in \mathbb{N}\}$$

(d) (10 points) Show that T is not regular (Hint: assume the opposite and use the argument above).

Solution: Assume T is a regular language. Using the closure property, $\bar{T}$ is also a regular language. Since a^*b^* is regular, then $a^*b^* \cap \bar{T} = a^*b^* - T$ is also regular.
But $a^*b^* - T = \{a^i b^i \mid i \in \mathbb{N}\}$ is not regular. Contradiction. T cannot be regular.

4. Let G be a context-free grammar (CFG):

$$\begin{aligned} S &\rightarrow SP \mid aPb \mid aQ \\ P &\rightarrow aPb \mid Q \mid \epsilon \\ Q &\rightarrow aQ \\ R &\rightarrow RQ \mid b \end{aligned}$$

(a) (10 points) Simplify the grammar (i.e., eliminate ϵ productions, unit productions and useless symbols, all in the correct order). Then, write the grammar under its Chomsky Normal Form (CNF). Show your work and justify each steps.

Solution: First, we eliminate ϵ-productions. Only variable P is a nullable symbol.

$$\begin{aligned} S &\rightarrow SP \mid aPb \mid ab \mid aQ \\ P &\rightarrow aPb \mid ab \mid Q \\ Q &\rightarrow aQ \\ R &\rightarrow RQ \mid b \end{aligned}$$

Then, we remove the (only) unit productions $P \to Q$.

$$\begin{aligned} S &\to SP \mid aPb \mid aQb \mid ab \mid aQ \\ P &\to aPb \mid ab \\ Q &\to aQ \\ R &\to RQ \mid b \end{aligned}$$

Next, we eliminate the variable Q that derives no string. To this end, we identify S and P as the only variable that derive strings of terminals. Therefore, Q is useless and should be removed.

$$\begin{aligned} S &\to SP \mid aPb \mid ab \\ P &\to aPb \mid ab \\ R &\to RQ \mid b \end{aligned}$$

Finally, we must validate that all variables are accessible from the start symbol S. We find that P is accessible but not R. Therefore, we remove the productions originating from R.

$$\begin{aligned} S &\to SP \mid aPb \mid ab \\ P &\to aPb \mid ab \end{aligned}$$

Now, let's build the CNF. We introduce first variables A and B producing the literals a and b, and use them to substitute a and b in the productions is the lefthand side has more than one symbol.

$$\begin{aligned} S &\to SP \mid APB \mid AB \\ P &\to APB \mid AB \\ A &\to a \\ B &\to b \end{aligned}$$

Finally, we add variables X, Y to build productions rules with only two variables in the right-hand side of the productions.

$$\begin{aligned} S &\to SP \mid AX \mid AB \\ X &\to PB \\ P &\to AY \mid AB \\ Y &\to PB \\ A &\to a \\ B &\to b \end{aligned}$$

Here, we can note that the new productions originating from X and Y and we merge them to simplify the grammar (this is optional but best).

$$
\begin{aligned}
S &\rightarrow SP \mid AX \mid AB \\
X &\rightarrow PB \\
P &\rightarrow AX \mid AB \\
A &\rightarrow a \\
B &\rightarrow b
\end{aligned}
$$

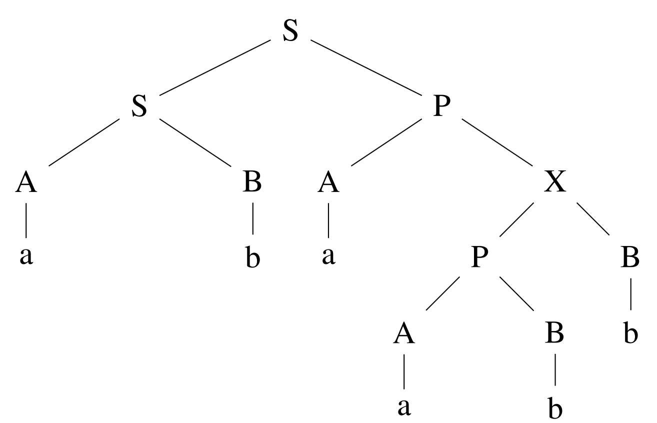
(b) (10 points) Use the Cocke-Younger-Kasami (CYK) algorithm to decide if the string $s = abaabb$ is a word of the language $L(G)$.

Solution:

$$
\begin{array}{cccccc}
\{\mathbf{S}\} & & & & & \\
\emptyset & \emptyset & & & & \\
\emptyset & \emptyset & \{S, \mathbf{P}\} & & & \\
\emptyset & \emptyset & \emptyset & \{\mathbf{X}\} & & \\
\{\mathbf{S}, P\} & \emptyset & \emptyset & \{S, \mathbf{P}\} & \emptyset & \\
\{\mathbf{A}\} & \{\mathbf{B}\} & \{\mathbf{A}\} & \{\mathbf{A}\} & \{\mathbf{B}\} & \{\mathbf{B}\} \\
\hline
a & b & a & a & b & b
\end{array}
$$

(c) (10 points) Draw a derivation tree for s computed from your execution of the CYK algorithm. Highlight the symbols used in the CYK table.

Solution: The start symbol S is found in the top cell and therefore $s \in L(G)$. The derivation tree is:



(d) (5 points (bonus)) Describe the language $L(G)$.

Solution: The set of correctly nested parenthesis (i.e., like "((()))()(())").

Question:	1	2	3	4	Total
Points:	20	10	40	30	100
Bonus Points:	0	0	0	5	5
Score:					