

Assignment 1

CMSPC 461: Programming Language Concepts

Prof. Suman Saha, Spring 2023

Due: 11:59 PM, - 1st Feb, 2023

Instructions:

You need to submit your homework to **Gradescope**. Do **every problem on a separate page and mark** them before submitting. **If the questions are not marked** or are submitted with **incorrect page-to-question mapping**, the question will **not be graded**. Make sure your name and PSU ID is legible on the first page of your assignment.

We highly recommend you follow the latex/doc template (which can be found on canvas) for the homework submissions, however, if you are writing and scanning make sure you maintain proper handwriting and use stationary of high contrast.

**(Kindly refer syllabus for late submission and academic integration policies.)

Question 1

(7*3 = 21 pt)

Construct regular expressions for the below 'English' statements. If you cannot construct one, explain why that may be the case.

1. Given an alphabet $\Sigma = \{a, b\}$, a string where a is repeated even times and b odd times.
2. Given an alphabet $\Sigma = \{a, b\}$, a string which does not end with aa.
3. Given an alphabet $\Sigma = \{a, b\}$, a string which does not contain aa.
4. Given an alphabet $\Sigma = \{a, b, c\}$, a string that has a symbol in the middle that is neither its start nor its end symbol, and its end and start symbols are different, for eg, abbbbbc or bcccca.
5. Given an alphabet $\Sigma = \{a, b\}$, a string with a's followed by b's with both the number of a's and b's being equal.
6. A general email ID with between 13 to 18 characters before the @ symbol and 4 characters for a domain name and 'edu' for a top-level domain.
7. Given an alphabet with opening and closing parentheses only, express a string that contains the sub-string $((())$.

Question 2

(6*3 = 18 points)

1. Read this interesting article on Java (<https://www.ibm.com/cloud/blog/jvm-vs-jre-vs-jdk>) and tell us if you think Java is interpreted or compiled and if both, when it is interpreted and when it is compiled.
2. Build a DFA to accept hexadecimal numbers. If a DFA has all its arrows inverted except the 'starting' arrow, would it still be a DFA?
3. Convert the above to an NFA
4. Generate a DFA for a set of strings with a's followed by b's with both the number of a's and b's being equal for the alphabet $\Sigma = \{a, b\}$
5. Generate a DFA for a regular expression defined for the language over an alphabet $\Sigma = \{a, b\}$ with only one b.
6. Construct an FA for the regex $a^*b^*a^*a$ with only 3 states

Question 3: Grammar

(3*5 = 15 pt)

1. Construct grammar for the below 'English' statement. Given an alphabet $\Sigma = \{a, b\}$, generate a grammar that represents strings that start with a when it ends with b and start with b when it ends with a.
2. Construct grammar for the below 'English' statement. Given an alphabet $\Sigma = \{a, b\}$, generate a grammar that represents strings with an equal number of a's and b's
3. Is the following BNF grammar ambiguous? If so, design an unambiguous grammar that produces the same binary-string language.
 $\langle \text{binary} - \text{string} \rangle \rightarrow 0 \mid 1 \mid \langle \text{binary} - \text{string} \rangle \langle \text{binary} - \text{string} \rangle$

Question 4

(5*3 = 15 points)

Note that the questions below have a sequence, so working through them in increasing order helps you firm up your concepts better.

G is a grammar of the form $S \rightarrow D | S - S | S + S$ where D represents mathematical digits.

1. Derive $6 - 3 - 4$ using leftmost derivation if it is possible.
2. Draw all possible parse trees for the above string.
3. Derive $9 - 1 -$ using rightmost derivation if it is possible.
4. Derive $8 + 6 + 3 - 5$ using a derivation that is not leftmost or rightmost (Note - we DO NOT want to see parse trees)
5. How many left most and right most derivations do we have for the above string? Why?

Question 5

(5*3 = 15 points)

The following is based on Question 4

1. How many parse trees can we construct for the string in Q4.4?
2. Is the number of derivations for the string in Q4.4 the same as the number of parse trees? Why/ why not?
3. Is this grammar ambiguous? Why/why not?
4. If the grammar is ambiguous (it need not be), try to design a new grammar to remove its ambiguity.
5. For the above question is it possible to do it in multiple ways, and if yes, in each such method, is there some 'meaning' to $+$ and $-$, or their order?

Question 6

(4*4= 16 points)

These questions are slightly tougher and may require to think deeply. Please attend office hours- we are here to help!

1. Consider the language consisting of strings that represent the list of numbers separated by commas. For instance, the string "10, 7" and "1, 7, 5, 13" are in the language; also included in the language are lists of a single number (e.g., "12"). Write an unambiguous BNF grammar for the language. Briefly explain why your grammar is unambiguous.
 2. $S \rightarrow aS | aSb | b | \epsilon$ Is this grammar ambiguous? Why? Can we remove its ambiguity if it is ambiguous (provide a replacement grammar)?
 3. The following grammar for arithmetic expressions allows addition, subtraction, as well as a unary operator "~" for negation; that is, "~ 8" is interpreted as the number negative eight.
$$\langle e \rangle \rightarrow \langle d \rangle \mid \langle e \rangle + \langle e \rangle \mid \langle e \rangle - \langle e \rangle \mid \sim \langle e \rangle$$
$$\langle d \rangle \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$$

The grammar is clearly ambiguous. Change the grammar so that "+" and "-" are left-associative and the precedence of "~" is higher than "+" and "-".
 4. Do context-free grammars represent all regular languages? And do regular languages translate all grammars? Are there any languages CFGs may not be able to represent? (Explain each answer in 10 words) (2*3 = 6 pt)
-