Guideline

Due Date: Thursday, 2023-09-14, by 23:59.

Upload your answers as a singular PDF to Brightspace.

Typewriting is preferred; If you're writing by hand, please ensure your handwriting is legible.

Multiple submissions are possible before the due time; the last submission will be graded.

Regular Expressions (points = 50)

Below are two abbreviations that are widely used in regular expressions:

- The expression `\d` is a shorthand for the regular expression [0-9].
- The quantifier `{n}` is used to match exactly n occurrences of the preceding character class. For example, the regular expression `\d{3}` matches exactly three consecutive digits in a string; the regular expression [a-z]{5} matches five consecutive lowercase letters.

Solve the following questions:

1. What does the regular expression `\d{3}-\d{2}-\d{4}` match?

   a) A phone number in the format (xxx) xxx-xxxx

   b) An email address

   c) A date in the format mm/dd/yyyy

   d) A Social Security number in the format xxx-xx-xxxx

2. What does the regular expression `[a-z]\d*` match?

   a) Any word containing only lowercase letters

   b) Any word containing only uppercase letters

   c) Any word containing only numbers

   d) Any word containing only a lowercase letter followed by optional digits

3. What does the regular expression `\d{3}-\d{3}-\d{4}|\d{10}` match?

   a) A Social Security number in the format xxx-xxx-xxxx or a 10-digit phone number

   b) A date or a phone number

   c) A phone number or an email address

   d) A Social Security number or an email address

4. What does the regular expression `[a-z]+?@[a-zA-Z_]+?.[a-zA-Z]{2,3}` match? Choose the closest answer.

   a) A phone number

   b) An email address

   c) A URL

   d) A street address

5. Please conduct some research. Your task is to determine the Linux command that can recursively search for all markdown files (with the ".md" extension) in the current directory that contains a negative integers. Note: to match the minus symbol, you can use `\-` or `[-]`. The following should be recognized: -89, -1, -007. The following should not be recognized: 0, -x, 42. A good starting point for your investigation might be familiarizing yourself with the `grep` command. You can refer to the Wikipedia page on grep for an overview.

Context-free Grammar (points = 50)

6. What is the language generated by the following grammar? `S -> aSb | ε`

   A. The set of all strings that with 'a' and end with `b`.

   B. The set of all strings that contain an equal number of 'a's and 'b's.

   C. The set of all strings that contain an even number of 'a's followed by an even number of 'b's.

   D. The set of all strings that contain n 'a's followed by m 'b's, where n = m >= 0

7. Create a grammar that generates all strings over {a, b} that start and end with the same symbol.

8. Given the grammar with the following productions:

   ```
   S -> aSbb | ε
   ```

   Determine the language generated by the grammar.

9. Given the following grammar

```
E -> E + E | E * E | id
```

Draw different parse trees for the string id + id * id to demonstrate ambiguity.

10. Given the following grammar

```
```

```
S -> aAb
A -> c | d
```

Can `acb`, `adb`, `adab`, `aab`, `ab` be parsed? Give an answer for each but you do not need to explain.

Column 1	Can be parsed (true/false)
acb	
adb	
adab	
aab	
ab	