COMP 221 Homework 4

Due: Friday, October 11, by 11:59 pm

1 Guidelines

Please **type up** and submit a PDF of your solutions. I recommend LaTex, but you are certainly welcome to use Google Docs or some other word processing program. Just make sure to save it as a PDF before submission. If you have to draw any pictures for a question (such as trees, graphs), you can do so **neatly** and then add a picture of it to your PDF file.

If a problem is explicitly marked as a **Programming** question, you will probably need to submit the source code for it (unless otherwise indicated). You should answer additional questions in the PDF.

Programming language: You are welcome to use Python or Java.

Do your own work: Homework should be individual work. You may discuss problems with other people, but you must write up your solutions yourself. I will not say that the web is off limits, but handing in solutions you find online as if they are your own is also not acceptable. Speaking of, most of these questions are shamelessly stolen from Susan :)

Make sure you have your name on your homework!

- **Question 1** Given the description, write the regular expression and draw the discrete finite automaton.

  1. A string of a's and b's with exactly two a's.

  2. A string of a's, b's, c's, and d's that accepts 0 or more strings of "ab" or 0 or more strings of "cd" in any order and in any quantity. This means that strings like "aba" would not be accepted, nor would "a", but "$abababab cdcdcdcd$" would be fine. "$cdcdababcdcdabab$" , "$abcdab$", and "$cdabcd$" would also be fine. The empty string would also be accepted.

  3. A string of a's and b's that has an even number of b's. The empty string would also be accepted.

- **Question 2**

  For the DFA on the next page, identify the start and accept states, and write 2 strings that would be accepted by the machine and 2 strings that would be rejected.

- **Question 3**

  Write a regular expressions for the DFA given on the next page.

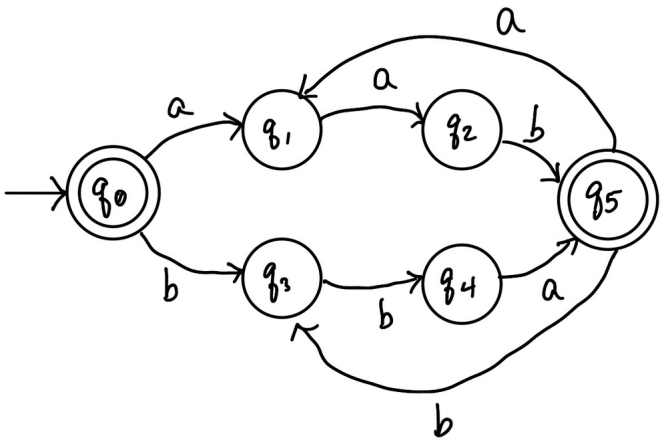

Figure 1: DFA for Questions 2 and 3

- **Question 4 - Programming**

  Use regular expressions for your preferred programming language to help you write a program that takes in a text file, and scans the text file for dates.

  Here are links to website that talk about the regular expression libraries for a number of programming languages.

  Python: https://docs.python.org/3/howto/regex.html

  Java: https://docs.oracle.com/javase/7/docs/api/java/util/regex/Pattern.html

  Look for dates in the following formats (allowing for any month, day, and reasonable (by your definition) year):

```
May 14, 2020   (Month Day, Year)
14 May, 2020   (Day Month, Year)
05/14/20      (MM/DD/YY, but you may opt to allow single-digit month and day)
05/14/2020  (MM/DD/YYYY, but again you may allow M/D/YYYY and similar)
2020/05/14    (YYYY/MM/DD)
05-14-2020   (MM-DD-2020)
```

  Your program should output all the dates it finds and matches. Test your program with an assortment of inputs to make sure it works. Remember that dates can be embedded in paragraphs of text as well.

  For your homework writeup, explain the different parts of your regular expression(s) and how they worked. Which ones were easy or hard to write? Are there date formats you can't recognize?

  Submit your code directly on Moodle.