BT 3051 — Data Structures and Algorithms for Biology

Jul–Nov 2019

Assignment 2

22nd October 2019

Due date: 30th October, 2019 @ 17:00   **Maximum marks: 35**

Instructions: Write regular expressions to solve the problems mentioned below. If you need any assistance, feel free to write to me or the TAs via Piazza (private note). Evaluation will be based on the codes and the logic.

Academic Integrity: You are allowed to discuss the problems verbally with your friends, but copying or looking at codes (either from your friend or the Web) is not permitted. Transgressions are easy to find, and will be reported to the "Sub-committee for the Discipline and Welfare of Students" and will be dealt with very strictly. Mention any collaboration (discussions only!) in your solutions.

Late submission penalties: 1 second – 24 h: 20%; 24–48 h: 40%; > 48h: 60%

Early submission bonuses: > 24h: 5%, > 48h: 10%, > 72h: 20%

Evaluation: Assignments will be evaluated by the TAs within two weeks of the due date. You can check out your marks and contest them, if needed, for at most one more week post-evaluation, i.e. three weeks from the due date of the assignment.

Problem Statement

1. (5 marks) Check if the following regular expressions match the target **exactly** from the given sample text. Give reasons in case there are no matches:
   i.   Target: 18
        Regex: **[129]|18**
   ii. Target: There are two waiters, two footmen, two bellboys, and an Arab teenager, small, cheerful, and alert, who appears to be some kind of page. He is Zero.
        Regex: **There.*Zero**

   Sample Text:
   A double-reception salon with high ceilings and two couches. There are 18 trunks and eight suitcases arranged neatly at the side of the room. Each is painted with the initials Mdm. C.V.D.u.T. Outside, a light snow fall ensues. A tall, blond, forty-year-old concierge stands patiently alone surveying the room. He is tranquil, perfectly composed, waiting. He wears the faintest hint of mascara. He is M. Gustave. M. Gustave crosses swiftly to the door and opens it just as a contingent of hotel staff arrives together from down the corridor. There are two waiters, two footmen, two bellboys, and an Arab teenager, small, cheerful, and alert, who appears to be some kind of page. He is Zero. One of the waiters carries a table, and one carries a breakfast tray.

2. (8 marks) **Match/Mismatch**

Write a regular expression for each of the below, such that it matches the first set of words and does not match the second set of words. Also, write one line explaining the regular expression. Carefully crafted regular expressions will get higher credit; trivial regular expressions will not get any credit (e.g. matching each word using '|'!).

(a) Match: Swell, Bell, Well, Intel
Mismatch: Felling, Filling, Farewell

(b) Match: Croquet, Quote, Quite, Joaquin
Mismatch: Lacquer, Acquire, Acquaint

(c) Match: Itinerant, Inimical, Initial, Tortellini
Mismatch: Intelligent, Inquisitive, Spaghetti

(d) Match: 12345, 23456, 78901, 7980, 5432
Mismatch: 12abc, 123, 578923, abc12

3. (10 marks) **Regular expression examples**
For each of the regular expressions given below, explain what each of them means and give examples of 4 words which **match and do not match** them. Underline the matches in the examples.

(a) \<a\w*e\w*i\w*o\w*u\w*\>

(b) \<\w*\w{2}\w*\w{2}\w*\>

(c) ^(\d{3})\d*(\1\d){1,}\$

(d) \w+\'\w+

(e) ^(((\d)\d)\d)\d\1\2\3\$

4. (12 marks) **Regular expressions**
Write regular expressions for each of the following:

(a) All real numbers (should not match any alphanumeric expressions)

(b) A valid date in *mm/dd/yyyy* format. Note that the expression should account for '_' and '.' as separators as well.

(c) Filename from a http/https address. For instance, in *https://some/address/directory/example.tar*, *example.tar* should be a match.

(d) The longest open reading frames in a given DNA sequence (ignore the complementary strand)

(e) The shortest open reading frames in a given RNA sequence (i.e. without internal stop codons)

(f) Passwords that meet the following criteria:

     i.   7-15 characters long

    ii.   At least one uppercase letter

   iii.   At least one lowercase letter

   iv.   At least one digit

    v.   At least one special character other than '*' and '@' (must not contain '*' and '@')

How to Submit your Homework

- Submit your assignment ONLY via the submission link: http://tinyurl.com/bt3051-submit.

- You should not be signed into Dropbox while uploading this file (or use an incognito window to open the link), so that you can enter the following details during submission, instead of Dropbox auto-filling it:

  - First Name: Roll Number
  - Last Name: Your Full Name
  - E-mail: Your `smail` id

- Save your solution files as `hw2.pdf` and submit the file. Do not use different filenames!

- Each of your submission files, `hw2.pdf` should begin with the usual **header information** — the number of the assignment, your roll number, your collaborators' roll number(s), and approximately how much time you took to solve the problems in that part of the assignment.

- **Submissions not adhering to any of the above instructions will not be evaluated**.

- Also do not send the files by e-mail — obviously, they will not be evaluated.