CS 301: Languages and Automata **Spring 2018**

Homework 2

Due: Sunday, February 18, 2018

Instructions

This assignment is due Sunday, February 18, 2018 at 11:59PM (Central Time). Solutions for Part I must be submitted on Blackboard and solutions for Part II must be submitted on Gradescope.

Late submissions will be accepted within 24 hours after the deadline with a penalty of 25% of the assignment grade. No late submissions will be accepted more than 24 hours after the deadline.

Part I: Regular expressions

Problem 1 Grep and sed. For this problem, you will write regular expressions in the POSIX extended regular expression format.[1] You will use these regular expressions with the provided `data.txt` file.

`data.txt` consists of a table of information regarding government offices including phone numbers of the administrators in charge of the office. Each line in the file consists of a number of tab-delimited fields. The first line of the file gives the name of each field.

a. **[5 points]** Grep. Write a POSIX extended regular expression that can be passed to `grep -E` to output all of the lines in `data.txt` containing phone numbers in the 202 and 314 area codes.

Save your regular expression in a one-line text file named `1a.txt`. Your regular expression will be tested by running

```
$ grep -E -f 1a.txt data.txt
```

so be sure to test your regular expression like this.

The first two lines of output of running that command should look like

```
N       111     Archives I      OFFICE OF THE ARCHIVIST OF THE UNITED STATES    202-357-5900    David S Ferriero
ND      4200    Archives II     DEPUTY ARCHIVIST OF THE UNITED STATES   202-357-5900    Debra Steidel Wall
```

b. **[15 points]** Sed. A sed substitution command looks like

[1] https://en.wikibooks.org/wiki/Regular_Expressions/POSIX-Extended_Regular_Expressions

```
s/regex/replacement/flags
```

The regular expression is matched against each line in the input and the part of the text that matches (if any) is replaced by the replacement. By using `\1`, `\2`, etc. in the replacement, sed will replace the string matching the regular expression with the replacement text but substituting the text matching a parenthesized subexpression for the `\1`. For example, running

```
$ sed -E 's/([^,]*),(.*)/\2:\1/' file.txt
```

on a file containing the line `foo,bar` will replace that line with `bar:foo`.

To only print the lines on which a substitution has been performed, we can pass the `-n` option to sed and use the `p` flag to the substitution command. E.g.,

```
$ sed -n -E 's/([^,]*),(.*)/\2:\1/p' file.txt
```

Your task is to write a sed substitution command that matches lines with phone numbers in the 202 and 314 area codes *and* has a nonblank person in charge. The substitution command should replace each matching line with

```
[Symbol] In Charge: Phone
```

where `Symbol`, `In Charge`, and `Phone` correspond to the fields in the line with those names.

Write the substitution command into a one-line text file named `1b.txt`. Your command will be tested by running

```
$ sed -n -E -f 1b.txt data.txt
```

so be sure to test that format. Note that since the `-n` option is being passed to sed, your command should use the `p` flag.

The first two lines of the output should be

```
[N] David S Ferriero: 202-357-5900
[ND] Debra Steidel Wall: 202-357-5900
```

Note that some of the lines with telephone numbers in the appropriate area codes do not have a person in charge and thus should *not* be printed. The line with the Symbol "AFN-MR" is an example of one that should not be printed.

Problem 2 [30 points] For each of the languages below, construct a regular expression in JFLAP, version 7, that generates each of the following languages. To receive full credit for each language, you must submit three files: (1) the JFLAP file (e.g., `2a.jff`); (2) a text file (e.g., `2a-accept.txt`) with five strings that are in the language, one per line; and (3) a text file (e.g., `2a-reject.txt`) with five strings that are *not* in the language, one per line. In total, you should have 30 files.

Note that JFLAP uses the `+` symbol rather than | or ∪. So the regular expression `aba` | `b`* should be written `aba+b*`.

Each language is worth 3 points.

a. $\{w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*$ and every pair of adjacent $\mathtt{a}$s is followed by $\mathtt{b}\}$

b. $\{w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*$ and every $\mathtt{a}$ in w must be preceded and followed by $\mathtt{b}\}$

c. $\{w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*$ has an odd number of $\mathtt{a}$s and ends with $\mathtt{b}\}$

d. $\{w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*$ has exactly one or two $\mathtt{b}$s$\}$

e. $\{w \mid w \in \{\mathtt{a}, \mathtt{b}\}^*$ starts with $\mathtt{a}$ and has at most one $\mathtt{b}\}$

f. $\{w \mid w \in \{\mathtt{0}, \mathtt{1}\}^*$ is a binary number that's a multiple of $3\} \cup \{\varepsilon\}$

g. $\{w \mid w \in \{\mathtt{0}, \mathtt{1}\}^*$ contains at least three $\mathtt{1}$s$\}$

h. $\{w \mid w \in \{\mathtt{0}, \mathtt{1}\}^*$ and $|w| \leqslant 5\}$

i. $\{w \mid w \in \{\mathtt{0}, \mathtt{1}\}^*$ and every odd position of w is $\mathtt{1}\}$

j. $\{w \mid w \in \{\mathtt{0}, \mathtt{1}\}^*$ contains an even number of $\mathtt{0}$s or exactly two $\mathtt{1}$s$\}$

Part II: Proofs

Remember, your solutions to Part II must be typeset. Handwritten solutions will not be graded and will receive a 0.

Problem 1 [5 points] Prove that the language $A = \{www \mid w \in \{\mathtt{a}, \mathtt{b}\}^*\}$ is not regular using the pumping lemma.

Problem 2 [5 points] Describe the error in the following "proof" that $B = \underline{\mathtt{0}^*\mathtt{1}^*}$ is not a regular language.

The proof is by contradiction. Assume that B is regular. Let p be the pumping length for B. Choose $w = \mathtt{0}^p\mathtt{1}^p$. We've seen several times that w cannot be pumped (E.g., Example 1.73 in Sipser). This is a contradiction so B must not be regular.

Problem 3 [20 points] Let $\Sigma = \{\mathtt{1}, \#\}$ and let

$$C = \{w \mid w = x_1 \# x_2 \# \cdots \# x_k \text{ for } k \geqslant 0, \text{ each } x_i \in \underline{\mathtt{1}^*}, \text{ and } x_i \neq x_j \text{ for each } i \neq j\}.$$

Prove that C is not regular using the pumping lemma. [*Hint: Select a string $w \in C$ such that w contains p # symbols.*]

Problem 4 Let

$$D = \{\mathtt{1}^k y \mid k \geqslant 1,\ y \in \{\mathtt{0}, \mathtt{1}\}^*, \text{ and } y \text{ contains at least } k\ \mathtt{1}\text{s}\}, \text{ and}$$
$$E = \{\mathtt{1}^k y \mid k \geqslant 1,\ y \in \{\mathtt{0}, \mathtt{1}\}^*, \text{ and } y \text{ contains at most } k\ \mathtt{1}\text{s}\}.$$

a. [10 points] Prove that D is a regular language.

b. [10 points] Prove that E is not a regular language.