MARIST

PROJECT | Formal Languages & Computability
CMPT 440
Spring 2015

For this project you will develop a version of the grep utility called graphex that searches files for regular expression pattern matches and produces dot graph file output for the automata used in the matching computation.

Name

```
graphex
```

Synopsis

```
java graphex.Grep [-n NFA-FILE] [-d DFA-FILE] REGEX FILE
```

Description

graphex should perform the following high-level steps.

- Learn the alphabet from the input ***FILE***.
- Convert the regular expression ***REGEX*** to a NFA.
- Convert the NFA to a DFA.
- Use DFA computation to test each line of the file for accept/reject.
    - File lines are delimited by newline characters
    - Accepted lines should be printed to standard output
- Optionally, output the NFA and/or DFA to the specified filenames in DOT language format.
    - Information on DOT can be found here:
        - http://en.wikipedia.org/wiki/DOT_language
        - http://www.graphviz.org/content/dot-language

Notes

- Each line should be considered for acceptance in its entirety. This is equivalent to using grep with start (^) and end ($) anchors surrounding the pattern.
- You may compute on the NFA if you prefer, but you still need to generate the DFA for DOT output.
- Your program should work with or without the DOT output files specified.

Submission

Save your Java files in a package named `graphex`

```
graphex/
    Grep.java
    ... any other .java files
```

Zip this file using `zip` to the file `graphex.zip`. Use **only** zip (not gzip, rar, etc). Submit this file through iLearn. Do **not** submit any other files.