Homework 5

Object Oriented Programming CSI 405

Problem

The Unix grep program finds lines that match a regular expression. Implement a java version of grep.

Make this implementation run faster by running it in a multithreaded way. Split each input file into two equal-sized regions, and look for instances of the regular expression in each region, in a separate thread. The second thread should not output any lines until the first thread is done; instead, it should simply record the matching lines internally, so that it can operate in parallel with the first thread, and wait until the first thread is done before outputting anything based on its records. If an input line straddles the boundary between the two regions, or begins at the very start of the second region, the first thread (and not the second thread) should output the line.

Measure the performance of the original Java grep (non-parallel), compared to your modified (parallel) version, and compare both to GNU grep.

If you have access to the ualbany Unix environment you may run the grep command as follows:

```
LC_ALL='C' time grep -n 'word to find' /path/to/file/abc.txt
```

This will print the timing for you

You may add timings directly in your java code in your main method (beginning and at end of scope)

Make sure your JAVA versions expect command line args in same format as the grep command does (shown above). For now, only the above type of args need to be handled

Grading

Implementation	50%
Execution	40%
Code clarity / organization	10%

Submission

Submit any specific instructions needed for grading/executing your homework.
Your project(s) should be a 7-zip (or any other zip format) file submitted on blackboard. It should be an eclipse project that the grader can import into his/her eclipse environment and execute. If this criterion is not followed you will lose points.

The UML diagrams (if told to create) should be submitted separately in the same submission as a folder of images well labelled/organized. (Not inside the project archive). Same for sequence diagram(s).
Follow exact deadlines, timing on blackboard. Rules for late submission apply as per syllabus.

Follow naming conventions posted on blackboard.