

This assignment may be done individually, or by a pair of students. Prepare your answers to the following questions in a plain ASCII text file. Submissions in any other format will not be graded. Submit your file to the Curator system by the posted deadline for this assignment. If you worked in a pair, only one of you should make a submission, but the file should contain the names and PIDs of both students. No late submissions will be accepted.

You will submit your answers to the Curator System (www.cs.vt.edu/curator) under the heading OOC07.

For questions 1-6, complete the grep command

```
grep -E _____ SomeFileName
```

by formulating an appropriate regular expression that is matched by exactly the results described in the question. When a string is referred to as a "word" in these questions, that refers to the way that grep determines "word" boundaries.

1. [10 points] The set of all 10-digit telephone numbers, like 540.231.5605, formatted in precisely that manner; match all and only telephone numbers that are preceded and followed by one or more spaces. (You do not have to take into account any restrictions on what are actually valid area codes or prefixes.)
2. [10 points] The set of all strings that consist of three or four lower-case letters, where the first character cannot be a vowel, and both ends of the string must be adjacent to space characters. (Be careful about the requirement that the string only contains letters.)
3. [20 points] The set of all lines in a file that begin and end with the word "the". (There are exactly 64 such lines in the Moby Dick file from the Gutenberg Project.)
4. [20 points] The set of all lines (in a file) that contain the place name "New England" or "Spain". (There are exactly 10 such lines in the Moby Dick file from the Gutenberg Project.)
5. [20 points] The set of all lines (in a file) that include a word that contains two (or more) consecutive occurrences of the letter 'a' or two (or more) consecutive occurrences of the letter 'u'. (There are exactly 6 such lines in the Moby Dick file from the Gutenberg Project.)
6. [20 points] The set of all lines (in a file) that contain the word "the" at least five times. (There are exactly 4 such lines in the Moby Dick file from the Gutenberg Project.)