6.045J/18.400J: Automata, Computability and Complexity Prof. Ron Rivest

Homework 2

Due: 5 March 2003

Problem 1: Prove that the unary language of perfect squares is non-regular. (That is, the language $L = \{1^n | n \text{ is a perfect square }\}$).

Problem 2: Let K and $L_2 \cdot K$ be regular languages. ($L_2 \cdot K$ is the concatenation of L_2 and K). Prove or Disprove the following statement: L_2 must be a regular language.

Problem 3: The pumping lemma states that if a language L is regular, then all sufficiently long strings in the language can be pumped. Prove that the converse of the pumping lemma is false. That is, prove that there exist non-regular languages which satisfy all the conditions of the pumping lemma.

Problem 4: In this problem, you will use the unix program grep, along with your knowledge of regular expressions, to find errors in a large text file. Today or Tomorrow, there will be a large text file posted on the course website, "6045-hw2file.txt" which has many errors (of the types described below). For each of the following errors, write a regular expression (using grep syntax) to identify the error and then use grep and your expression to find all the lines in "6045-hw2file.txt" that contain this error. For this problem you need to turn in a printout that contains the grep expression that you used and the output obtained by running grep on "6045-hw2file.txt". **Note:** For information about grep syntax and how to use grep, see the grep man page or the website, http://pegasus.rutgers.edu/ elflord/unix/grep.html

1. Find all occurrences of the word "teh" (A common misspelling of "the")
2. Find all words that begin with two capitol letter. (e.g. the word "EArth" or "GErmany")
3. Find all places where a period or comma is placed outside quotation marks instead of inside. (e.g. Find things like "Bob is cool", or "Bob can dance". If you want, you can also find cases where a colon or semicolon is placed inside the closing quotation mark)
4. Find all words that violate the rule, "i before e, except after c" (e.g. the word "releive" or "beleif". Of course, this rule only applies when the vowel sound is a long 'e'. You might consider specifically excluding common exceptions like "neighbor" or "weight" where the vowel sound is not a long 'e')
5. Find all places where the same word is repeated twice in a row. (This cannot be done with standard regular expressions, but the grep syntax allows problems like this to be solved quite easily)