Name: _____

Exercise C3, 3/22/05

Part 1: Language Identification

Regular expressions are one of the most commonly used way to specify regular languages. Since most modern computing environment support the use of regular expression, it would be good to practice specifying a regular language using a regular expression.

As an example, let us observe a very crude way to identify Japanese names from a list of names (http://www.tcnj.edu/~komagata/csc460/05s/misc/names.txt). The first approximation is to pick up the basic Japanese syllabic structure, which is a repetition of consonant-vowel sequence. A regular expression $(CV)^+$ captures the pattern, where $V = \{$**a**, **e**, **i**, **o**, **u**$\}$ and $C = LowerCaseLetters - V$. Using the Unix **egrep** utility, we can see how well this pattern picks up Japanese names.

```
% egrep -ni '^([bcdfghjklmnpqrstvwxyz][aeiou])+$'
   ~komagata/www/csc460/05s/misc/names.txt
83:Duda
142:Ho
150:Hu
166:Kane
177:Ko
197:Levi
198:Levine
199:Li
243:Pi
245:Pulito
250:Rice
259:Sami
299:Su
324:Vita
325:Vu
345:Wu
346:Xu
348:Ye
350:Yi
351:Yi
354:Yu
```

Since most names are not really Japanese (although you may not be able to tell), this pattern is not really good. But the accuracy can be improved by fine tuning the pattern.

Task: Pick one human language and create a single regular expression that would pick up the names in that language to some extent. Include the regular expression and a test run on the name list. Human languages are full of exceptions. So, limit your time and satisfy yourself at some point.

Note: If you prefer, you can use an alternative programming language/environment that supports regular expressions, e.g., perl, PHP.

Part 2: Distinguishing Regular and Non-Regular Languages

FSAs are extremely simple and fast automata. So, if your problem (set, language) is regular, you should definitely use an FSA, rather than a PDA or a TM as the abstract model for developing a program. Then, the question would be to analyze whether the problem is regular.

Task 1: Consider the following problem: Liisa realized that for any day, she receives m important (i.e., non-junk) e-mail messages followed by n junk messages, where $m + n = 2k$ for some integer $k \geq 1$. First, represent the problem as a set. Then, identify whether the problem, interpreted as a language, is regular or non-regular. Justify your answer at least to the level discussed in class.

Task 2: Consider the following problem: Mikko realized that whenever he wins the card game m consecutive times, he loses n consecutive times after that, where $n = 2m + 1$. First, represent the problem as a set. Then, identify whether the problem, interpreted as a language, is regular or non-regular. Justify your answer at least to the level discussed in class.

// End