Problem Set #4
Due: May 5th (in class)

Reading: Chapters 11, 43, and 61 of The New Turing Omnibus and lecture notes for the fourth week of class (“Exact Match” and “Indexing”). Write your answers in the space provided, or attach extra sheets as necessary.

1. How many comparisons does the simple string search (“naive search”) algorithm make in each of the following searches? Note: P = pattern, T = text, and the algorithm scans the text to see if the pattern occurs anywhere.

   (a) P = genome
       T = general info on genes and genomes

   (b) P = ATG
       T = ACCGTATCAATGAG

   (c) P = ruby
       T = web services with perl
2. Repeat the first problem, but this time count the number of comparisons the Boyer-Moore algorithm would make. Hint: write the pattern below the text, and show how far the “window” moves after each mismatch.

(a) P = genome
T = general info on genes and genomes

(b) P = ATG
T = ACCGTATCAATGAG

(c) P = ruby
T = web services with perl

3. Below is an FSA for the pattern ATCAT. What states does it go through as it searches for this pattern in the text TAGACTATCATTAT?
4. Draw an FSA for the pattern TATAA. At a minimum you should have the “success” transitions and failure transitions that are taken when the FSA finds a C or G in the text.

For extra credit include the transitions for mismatches with A or T that allow the machine to remember partial matches, e.g. your new machine should succeed on TTATAA.