TRA301/COS401: Homework 1
Due Date: February 19, 2009

1. **Lexicon Maintenance Tool:** Imagine you work at ManTra translation company. You are the person responsible for building and maintaining a lexicon of English words for the company. Implement a lexicon maintenance tool using the finite-state automata library that we have discussed in class that supports the following:

   (a) Read in a list of words from a file "1a" (The file of words will be on the course website) and constructs a finite-state acceptor (FSA) for that list. How many states and arcs does your FSA have (use fsminfo -n)?

   (b) Represent the FSA to take up the least amount of space (in bytes). How many states and arcs does your FSA have?

   (c) Using FSA operations, add the words in the file "1c" to your lexicon. Show the operations for doing this task. How many states and arcs does your FSA have?

   (d) Using FSA operations, delete all the words that start with a 'z' and end with an 's' from your lexicon. Show the operations for doing this task. How many states and arcs does your FSA have?

2. **Morphological Analyzer/Generator:** Your boss at ManTra comes to you with some "exciting" challenge (a euphemism for more work) without any pay raise. You groan, but given the current economic times you "gladly accept" the challenge, instead of quitting the company. Here is the challenge:

   (a) Implement a system that performs morphological analysis – takes in a word and produces its stem and its morphological features, using a finite-state transducer (FST). The words and their morphological features needed for your FST are available to you on our course website in the file "2a". How many states and arcs does your FST have?

   (b) Show the results of morphological analysis for the words: "banks", "cutting", "like", "means", "up", "wet", "buffalo" using your FST. Illustrate the FST operations you use for this task.

   (c) Use the same FST for morphological generation. Take as input the file "2c" on the course website and for each line containing morphological features use the FST you built in 2a to create the inflected of the word. Show how you solve this task and the resulting word forms.