CS 229, Fall 2003 **Homework #2**

This homework is due in class on Monday, November 10. It covers Chapter 3, Sections 3 and 4 You can work with other people in the class, but you should write up your solutions in your own words to turn in.

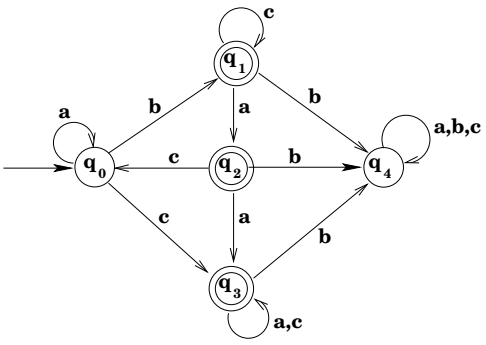
1. This problem is about the type of applied regular expressions that are discussed in Section 3.3, using an extended syntax.

a) Suppose you have a large text document and you want to make sure that it doesn't make the mistake of using "a" instead of "an" before a word that begins with a vowel (such as: "A elephant"). You want to search for such mistakes using a regular expression. Write a regular expression that will match a space, followed by an upper or lower case A, followed by another space, followed by a vowel (a, e, i , o, or u, upper or lower case). What cases of things like "a elephant" are missed by this regular expression? Write another pattern that matches the missed cases.

b) Shown below are some lines from a Linux passwd file. Each line contains 7 fields, separated by colons (:). The first field is a user name. The fifth field is the full name of the user. Suppose you have a large file consisting of lines like these, and you want to convert it into a file in which each line contains only a user name and a full name, separated by a comma. How could you do this using regular expression search and replace? What search pattern would you use? What replacement pattern would you use?

```
eck:x:501:102:David Eck:/home/eck:/bin/bash
vaughn:x:502:102:John Vauhgn:/home/vaughn:/bin/bash
cmc:x:503:102:Carol Critchlow:/home/cmc:/bin/bash
```

2. Consider the following DFA:



a) Based on this diagram, give a table for the transition function of this DFA.

b) Which of the following strings are accepted by this DFA?

a, b, abc, $aabaab$, $cccaaa$, $bbaabb$, $abac$, ε

3. Find a DFA that accepts each of the following languages. Give your answers as state diagrams. Briefly explain in English how each DFA works.

a) $L_1 = \{w \in \{a,b\}^* \mid w$ contains exactly three b's $\}$

b) $L_2 = \{w \in \{0,1\}^* \mid w$ begins and ends with the same symbol $\}$

c) $L_3 = \{w \in \{a,b\}^* \mid w$ contains the substring $abab$ $\}$

d) $L_4 = \{w \in \{a,b\}^* \mid$ the number of a's in w is odd and the number of b's is a multiple of 3 $\}$