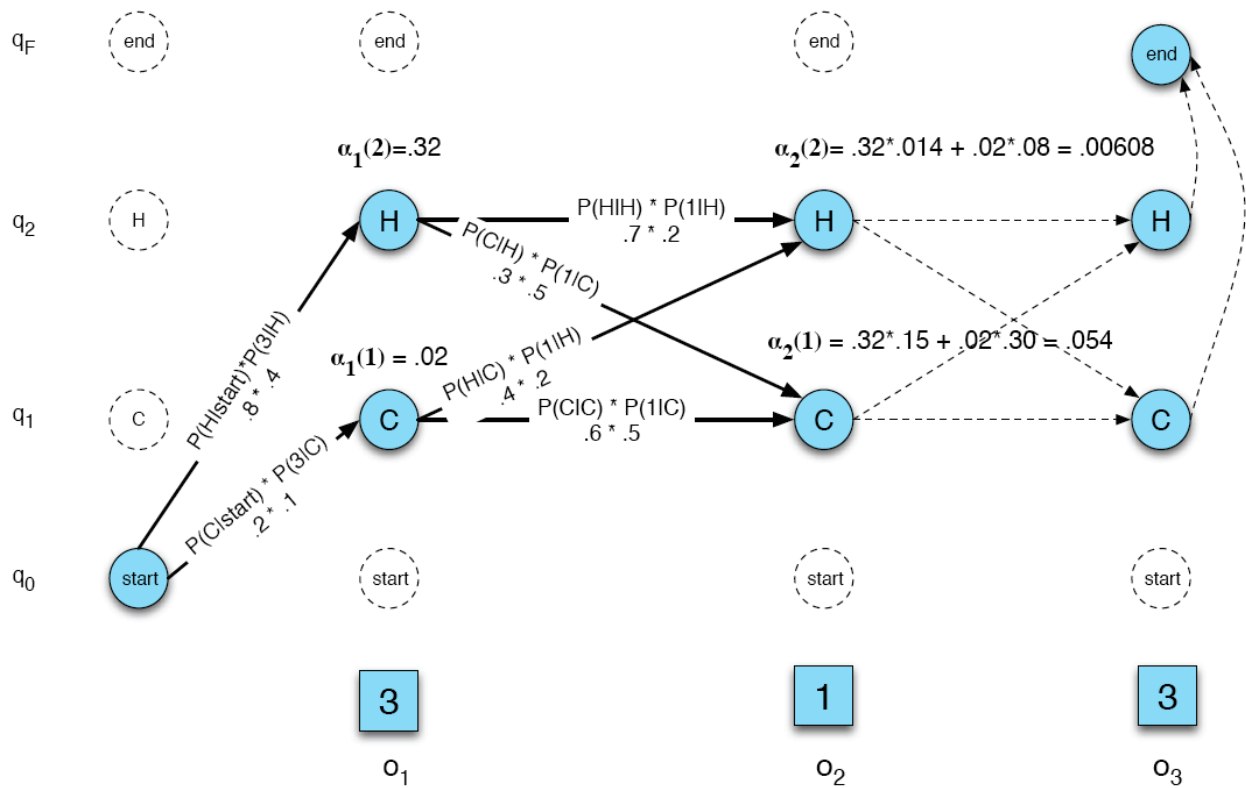CSCI 5832 Fall 2012 Exam 1

Name: _____

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. _____.

1. **(5 points)** What is your favorite movie?

2. **(5 points) True/False:** Given a finite alphabet, there are well-specified formal languages that **can not** be recognized by regular expressions, FSAs or regular grammars.

3. English has a class of adjectives that make use of the –ed suffix, as in "powdered milk", "boxed set", and "candied apple".
    a) **(5 points)** Is the process of going from the stem form to the –ed form in these examples an instance of inflectional or derivational morphology?

    b) **(10 points)** Describe the processes that would be involved in morphological processing of the "candied" example using FSTs.

4. Our sheep talk language is succinctly captured by the regular expression (baa*!).
    a) **(10 points)** Given access to a corpus of utterances by native speakers of this language, describe how you would *instead* capture this language as statistical language model using character-level bigrams. Include all the steps that would be involved.

    b) **(5 points)** What capabilities of the regex approach are lost in your new probabilistic approach (if any?).

5. **(10 points)** Parts of speech (or lexical categories) for words are determined by their morphological behavior and syntactic distributional properties. Give examples of each of these (morphological and syntactic) for the noun *school.*

6. Consider the Ice Cream HMM shown below. Using the notation and variables from this figure characterize the following values (symbols and variables are fine; you don't need to compute the actual answers):

   a) **(5 points)** The a priori probability of the sequence HHC (a priori means without evidence; that is without any particular observation sequence).

   b) **(5 points)** The probability of the state sequence CHC given the observation 313.

   c) **(5 points)** The probability of being in state H at time 3, given the observation 313.

   d) **(5 points)** Is the Viterbi algorithm of any use in computing the answer to part b?

In this figure, time starts at 1; Cold states have the value 1, and Hot states have the value 2. So the notation \alpha_i(j) represents the \alpha value of state j at time i.