Quiz 2 COMP 423 March 8, 2006

Answer all questions in the exam book. You may keep this exam sheet.

Calculators or other electronic devices are NOT permitted.

The Quiz is marked out of 15.

1. **(3 points)**

   (a) What is the "inverted file" of the following: `abccbaabc` ?

   (b) How can "header" information be used to compress inverted files, assuming gap lengths are encoded?

2. **(4 points)**

   (a) Encode the following sequence using LZ2 (sliding window):

   `baabaaaabb`

   Assume the window size is $n_w = 4$.
   You answer should include a parsing of the above sequence into phrases.

   (b) For the same sequence as in (a), show the phrase table built by an LZ3 decoder.

3. **(3 points)**

   (a) If the alphabet is $\{0, 1\}$, then a Lempel Ziv encoder takes any finite sequence of bits (a sequence) and maps it to another finite sequence of bits (a codeword for the sequence). Thus, the LZ algorithm defines a code on the set of finite bit strings. This code is not a prefix code. Why not? Give a counter example (and specify the LZ method).

   (b) How could one modify the LZ code so that it is a prefix code?

4. **(2 points)**

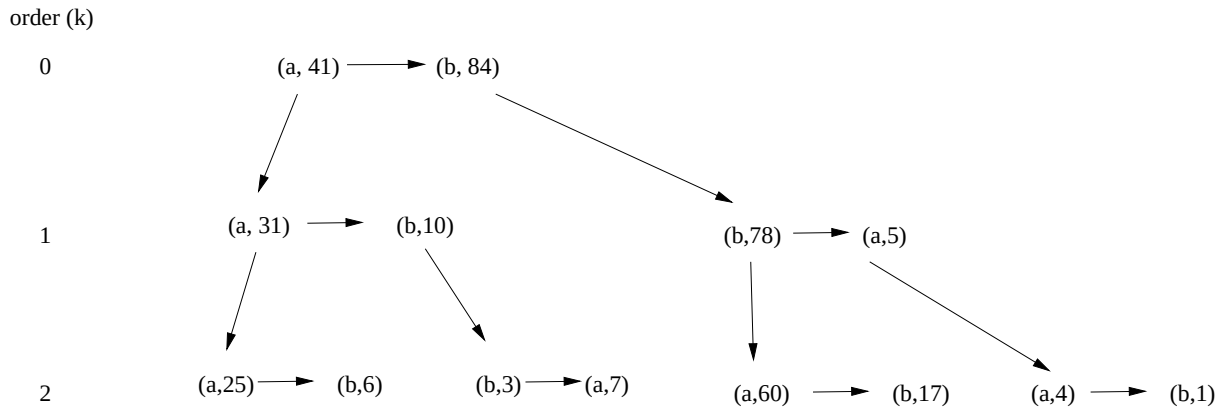
Consider the following conditional probability matrix for a first order Markov model, where the number of symbols in the alphabet is $N = 3$. Assume that these conditional probabilities are the same for all j.

$$P(X_{j+1} \mid X_j) = \frac{1}{8} \begin{bmatrix} 1 & 3 & 2 \\ 1 & 4 & 1 \\ 6 & 1 & 5 \end{bmatrix}$$

If the marginal probability $p(X_1)$ is uniform, then what are the marginal probabilities $p(X_2)$ and $p(X_3)$?

5. **(3 points)**

The following trie shows the frequency counts of a sequence, after j symbols have been seen by the encoder.



(a) Estimate the conditional probability of X_{j+1}, assuming:

- a 0^{th} order Markov model;
- a 1^{st} order Markov model;

You may assume that the alphabet is $\{\mathtt{a}, \mathtt{b}\}$ only. Be sure to state any other assumptions you make.

Hint: the count at a node at level $k+1$ is the number of times that the symbol at that node is the next symbol, given the previous k symbols.

(b) What were the symbols X_j and X_{j-1}?