CSC 310, Spring 2002 — Assignment #1

Due at **start** of tutorial on February 8. Worth 5% of the course grade.

Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

Question 1 (30 marks): For each code shown below, answer the following questions:

1) Do the lengths of codewords in the code satisfy the Kraft-McMillan inequality?
2) Is the code instantaneously decodable? If you answer no, give an example of a codeword that is a prefix of another codeword.
3) Is the code uniquely decodable? If you answer yes, show why this is so. If you answer no, give an example of a sequence of code symbols that can be decoded in two different ways.

Here are the codes, each of which is given as a set of codewords. The first uses the code alphabet $\{x, y, z\}$; the others use the binary ($\{0, 1\}$) code alphabet.

A) $\{xx,\ xz,\ y,\ zz,\ xyz\}$

B) {000, 10, 00, 11}

C) {100, 101, 0, 11}

D) {01, 100, 011, 00, 111, 1010, 1011, 1101}

E) {01, 111, 011, 00, 010, 110}

Question 2 (30 marks): Consider a binary Huffman code for a source alphabet with 2^k symbols, where k is a positive integer. Suppose that the probabilities for these symbols are such that $p_i/p_j < 2$ for all i and j in $\{1, \ldots, 2^k\}$. Prove that all the codewords in the Huffman code are k bits long.

Question 3 (40 marks total): Consider a source with source alphabet $\{W, B\}$, in which the probability of W is 0.99 and the probability of B is 0.01. Assume the source symbols are independent of each other.

The n-th extension of this source has as its source alphabet the set of all possible blocks of n W and B symbols, with probabilities found by multiplying the probabilities of the W and B symbols in the block.

Part A (15 marks): For $n = 2$, list all possible blocks and their probabilities, find and display (as a codeword table) a Huffman code for this extension, and compute the average codeword length for this Huffman code.

Part B (15 marks): Do the same for $n = 3$.

Part C (10 marks): Compute the entropy of the original source, and compare with the average codeword length divided by n found in Parts A and B. Show (without any tedious calculations) that getting the average number of bits per source symbol down to 10% above the entropy (or less) will require using blocks of size at least $n = 12$.