Algorithms in the Real World
Assignment #1 Due Date: 28 Sept 99

Collaboration policy: You can work in groups of two on this assignment.

Problem 1: (10pt) Exercises.

(a) Given the following conditional probabilities for a two state Markov Chain what factor would one save by using the conditional entropy instead of the unconditional entropy?

$$\begin{array}{ll} p(w|w) = .9 & p(b|w) = .1 \\ p(w|b) = .4 & p(b|b) = .6 \end{array}$$

(b) Would the following sequence of values

```
10 11 12 11 12 13 12 11 10 -10 8 -7 8 -8 7 -7
```

compress better using a cosine transform across the whole group of 16 or by blocking the 16 entries into two groups of 8 and transforming each separately? Explain briefly? (Assume equal quality.)

Problem 2: (30pt) Prove the Kraft-McMillan inequality for Prefix Codes. Note that this is somewhat easier than for general uniquely decodable codes.
You will be graded on the clarity as well as the correctness of your proof. Remember that making the proof longer does not necessarily increase the information content.

Problem 3: (60pt) Implement a compression algorithm, that works on byte files. You will be graded based on how well it compresses the Calgary Corpus. The grade will be based on the following equation:

$$\text{Points} = 60 - 30(B - 2)$$

where B is the average bits per byte for your algorithm. Up to 15 additional points will be given for originality of the algorithm.
You must hand in both the code, and a one page description of how the compression works (postscript). The source code and description should all be contained in a single directory, and the description should include the names of both partners. When you are finished you should copy the directory into `/afs/cs/academic/class/15853-f99/handin1/`. Please remove any old object or test files before copying so that you do not fill the class disk quota. Your encoder and decoder should take a single file from stdin and output to stdout. You can use any programming language as long as you include a Makefile that will compile the two programs `encode` and `decode` by running `make encode` and `make decode` (I can make exceptions to this rule if you ask).
You can use any existing code within your code, under the following conditions

1. The code you use is not proprietary.

2. You cite any code you use both in the header of the source file in which it is used, and in the 1 page description.

3. The code you use must not consist of more than 1/2 of your total code.