CS 337 **Quiz 3** **5/8/03**

Open book and notes.

Max points = 50 Time = 50 min Do all questions.

1. (Compresion; 19 points)

   (a) (4 points) Compute the entropy of the alphabet $[a, b, c, d]$ with associated probabilities $[1/2, 1/8, 1/4, 1/8]$.

   (b) (7 points) Let T be the tree corresponding to an optimal prefix code. Show that every non-leaf node in T has two sons. You may assume that every symbol has non-zero probability of occurrence.

   (c) (8 points) Let R be a set of symbols which includes two specific symbols, x and y. Let

   $$S = R - \{x, y\} \cup \{z\}$$

   where z is a new symbol not in R. Use x, y and z also for the probabilities associated with these symbols, and suppose $z = x + y$. Let r and s be the weights of the optimal trees of R and S, respectively. How are r and s related?

2. (Powerlist, 18 points)

   (a) (9 points) For each natural number i, u_i and v_i are powerlists, defined as follows.

   $$u_0 = \langle 0 \rangle, \quad v_0 = \langle 1 \rangle$$
   For all i, $i \geq 0$:
   $u_{i+1} = u_i \mid v_i$, and
   $v_{i+1} = v_i \mid u_i$

   Show that for all i, $i \geq 0$:

   $u_{i+1} = u_i \bowtie v_i$, and
   $v_{i+1} = v_i \bowtie u_i$

   Hint: You may prove only one part, say, $u_{i+1} = u_i \bowtie v_i$; the other proof is symmetric.

   (b) (9 points) Define the prefix function over powerlists. Write $p \sqsubseteq q$ to denote that powerlist p is a prefix of powerlist q. Use a pseudo programming notation, similar to the one used in the class for powerlists.

3. (String Matching; 13 points)

   (a) (Core computation; 6 points) You are given a string $v[0..20]$ and told that $v[6] \neq v[11]$. Which prefix of v can not be its core?

   (b) (Core computation; 7 points) Prove or disprove: for non-empty u and v,

   $$u \preceq v \;\Rightarrow\; c(u) \preceq c(v)$$