T-106.6200
Data Compression Exam

Instructor: Travis Gagie

May 17th, 2011
9–12 in T1

No written material is allowed in this exam. Submit at least one answer sheet, even if an empty one! Write on *each* answer sheet you submit the code of the course, the date, your name, and your student ID number. Each of the following questions is worth 1 point:

1. BUILD-A-CANONICAL-HUFFMAN-TREE-AGAIN

   (Include "-" as a character.)

2. Give the formula for $H_0(S)$.

3. Compare prefix-free coding, alphabetic coding and LZ77/78.

4. Compare the 0th-order, 1st-order and 2nd-order empirical entropies, i.e., $H_0(S)$, $H_1(S)$ and $H_2(S)$. As an example, consider $S = $ mississippi.

5. Compute the Burrows-Wheeler Transform of mississippi\$, considering \$ to be lexicographically smaller than the other characters and contexts to be either forward or backward as you choose.

6. Explain how to invert the transformation you just computed.

7. Occam's Razor is often stated as "given two theories with equal explanatory power, the simpler one is more likely to be correct". Your eccentric neighbour explains everything unexpected by exclaiming "*Ghosts!*" and claims that, since his theory has only seven characters, he's almost certainly right. Why is he wrong? (Hint: Refer to the more precise version of Occam's Razor that we discussed in the course — i.e., "we should choose the model so as to minimize the combined length the encoding of the model and the encoding of the data with respect to the model".)