

# T-106.6200

## Data Compression Exam

Instructor: Travis Gagie

May 10th, 2011  
16.15–18 in T2

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-FOR-THIS-SENTENCE  
(Include “-” as a character.)
2. State either the Kraft Inequality or Jensen’s Inequality.
3. Compare Elias’ gamma and delta codes. Give examples.
4. Compare LZ77 and LZ78. Give examples.
5. Describe each of the steps in the compressor  $\text{HUFF}(\text{RLE}(\text{MTF}(\text{BWT}(\text{file}))))$  individually, then say why they work well together.
6. Write a few sentences about the relationship between the  $k$ th-order empirical entropy  $H_k(S)$  of a string  $S$  and the number of bits needed to store  $S$ . Remember to consider the case when  $H_k(S) = 0$ . For example, if  $H_5(S) = 0$  and I give you a list of all the sextuples in  $S$ , can you reconstruct  $S$ ?
7. Alice and Bob want a list of ten thousand randomly chosen bits for an experiment, so Bob pays Charlie to flip a coin ten thousand times and write down the results. When Bob shows Charlie’s list to Alice, she claims Charlie cheated — his bits are *the 101st to 10100th bits of the binary expansion of  $\pi$* ! Embarrassed, Bob objects that she can’t be sure Charlie cheated because “one must not verify an idea using the same data that suggested the idea in the first place”. Alice counters that, by Kolmogorov complexity theory, the probability that Charlie’s list was truly randomly generated is less than  $1/2^{9559}$ . First explain in words why Alice is so suspicious of Charlie’s list, then show where she got that number. (Hint: There are 55 ASCII characters in the italicized text.)