Data Compression Techniques

Separate Exam, 14 September 2012 at 16-20 Lecturer: Juha Kärkkäinen

Please write on each sheet: your name, student number or identity number, signature, course name, exam date and sheet number. You can answer in English, Finnish or Swedish.

1. [*3+3+3+3 points*] Define the following concepts:

   (a) semiadaptive model

   (b) compression boosting

   What is the *main connection* between the following pairs:

   (c) wavelet tree and compressed permutation

   (d) Burrows–Wheeler transform and move-to-front encoding

   A few lines for each part is sufficient.

2. [*6+6 points*] Let $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}, \mathtt{g}, \mathtt{h}\}$ be the alphabet with the probability distribution

symbol	a	b	c	d	e	f	g	h
probability	0.10	0.15	0.15	0.05	0.25	0.10	0.15	0.05

   (a) Construct a Huffman code for the above symbol distribution.

   (b) Show that the code of (a)-part satisfies Kraft's inequality with equality.

3. [*12 points*] Describe Shannon's Noiseless Coding Theorem and discuss its significance for data compression. Your answer should cover both the lower bound and the upper bound part of the theorem.

4. [*6+3+3 points*] Consider a black-and-white image of $n \times n$ pixels with m black pixels, where m is much smaller than $n^2/2$. One method for compressing such an image is to encode the number of white pixels in-between each pair of consecutive black pixels (in row-by-row order, for example) using the gamma code.

   (a) What is the asymptotic size of the compressed image in the worst case (poorest compression) for a given n and m? *Hint:* The worst case arises when the black pixels are evenly distributed.

   (b) What is the asymptotic size in the *best* case?

   (c) Would using the delta code instead of gamma code change the asymptotic size in either the worst case or the best case?

   Justify all your answers.

5. [*6+6 points*] Encode the text `woodchuck␣could␣and␣would␣chuck␣wood` using

   (a) LZ77

   (b) Re-Pair.