Data Compression Techniques

Separate Exam, 21 August 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) move-to-front encoding

   (b) backward search

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

   (a) uniquely decodable code and Kraft's inequality

   (b) rank and select

   A few lines for each part is sufficient.

2. [*12 points*] Recall that

$$\gamma(n) = \text{unary}(k) \cdot \text{binary}_k(n - 2^k + 1) \qquad \text{where } k = \lfloor \log(n+1) \rfloor$$
$$\text{GR}_2(n) = \text{unary}(q) \cdot \text{binary}_2(n - q4) \qquad \text{where } q = \lfloor n/4 \rfloor.$$

   Give all integer ranges such that

   (a) the gamma code is shorter that the Golomb–Rice code GR_2

   (b) the gamma code is longer that the Golomb–Rice code GR_2

   (c) the gamma code and the Golomb–Rice code GR_2 have the same length.

   For each range, give the gamma and Golomb–Rice codes for the first integer in that range.

3. [*12 points*] Discuss how *adaptive* arithmetic coding differs from static arithmetic coding. Be as complete as possible. You may assume that the reader is familiar with static arithmetic coding.

4. [*6+6 points*] Compare LZ77 and LZFG (both without distance or length limits).

   (a) Give an example of a text, where the number of phrases in the LZFG parsing is significantly larger than in the LZ77 parsing.

   (b) Prove that the number of phrases in the LZFG parsing is never smaller than the number of phrases in the LZ77 parsing.

5. [*6+6 points*] Let $T =$ `monotonetone$`.

   (a) Compute the Burrows–Wheeler transform (BWT) of T.

   (b) Give a Huffman wavelet tree of the BWT of T.

   Give enough intermediate steps to show how you arrived at the solutions.