Chair for Algorithms
and Data Structures
Prof. Dr. Hannah Bast
Elmar Haußmann

Information Retrieval
WS 2015/2016

http://ad-wiki.informatik.uni-freiburg.de/teaching

Exercise Sheet 4

Submit until Tuesday, November 17 at **2:00pm**

Exercise 1 (5 points)

Prove that the Elias-Delta code is prefix-free (2 points) and specify the *exact* length of the code for integer x (3 points).

Exercise 2 (5 points)

Prove that there is no prefix-free code with length $\log_2 x + O(1)$ for the code for integer x. Hint: use part (1) of the central lemma from the lecture.

Optional: Prove the same for the length $\log_2 x + \log_2 \log_2 x + O(1)$.

Exercise 3 (5 points)

In the lecture, we derived that the probability distribution for a fixed gap of an inverted list is $p_i = (1-p)^{i-1} \cdot p$ for $i = 1, 2, \ldots$, for some p with $0 < p < 1$.

Verify that this is indeed a probability distribution over the natural numbers (1 point).

Then prove that the Golomb code is entropy-optimal for this distribution and for which value of M (4 points). Hint: try $M = c \cdot 1/p$, for a suitable c, and use that $1 - p \leq e^{-p}$. It suffices if you prove $L_i \leq \log_2 1/p_i + O(1)$ instead of the more strict $L_i \leq \log_2 1/p_i + 1$. Feel free to also try the latter, but it is much harder.

Exercise 4 (5 points)

Let $p_1, \ldots, p_n$ be the probabilities of a fixed distribution and let $L_1, \ldots, L_n$ be variable code lengths with $\sum_{i=1}^{n} 2^{-L_i} \leq 1$. Show that under this constraint $\sum_{i=1}^{n} p_i \cdot L_i$ reaches its minimum when $L_i = \log_2(1/p_i)$. Hint: Use Lagrange optimization, as explained on slide 19 of Lecture 3, and as briefly sketched on slide 21 of Lecture 4.

Commit your solutions in a single PDF in a new sub-directory *sheet-04* of your folder in the course SVN, and commit it. We recommend that you typeset your solution using LaTeX. With careful handwriting, you may also hand in a scan; in that case, take care that the scan has sufficient resolution and the file is not too large (< 1 MB). Also commit the usual *experiences.txt* where you briefly describe your experiences, your favorite movie and what you had for breakfast.