

Homework Assignment 3

Due date: 25th of November 2013

Please note that even though the homework assignments are optional, you're still highly encouraged to answer them, as they will help you prepare for your final exam. You can work in a group of two or alone. Solutions can be dropped off at the institute's homework mailbox located on the 2nd floor, next to room 238. In that case, please make sure both your name and matriculation number is noted down. If your answers span more than one sheet, kindly staple them together. Another alternative is to send your homework via email to: elmaarry@ifis.cs.tu-bs.de

LECTURE 4: INDEXING

EXERCISE 4.1

Referring to the document collection discussed in the fourth lecture: Indexing (slide 38). Compute the number of bits that are required to store the eighth index term, namely the term *old* for the following index representations:

- i. Fixed-bit code
- ii. Elias' gamma bit code
- iii. Elias gamma bit code with stored gaps.

EXERCISE 4.2

To make your index efficient, the number of disk accesses should be kept minimal, how can this be attained? Apply the proposed operation on the index representation you built in Exercise 4.1.iii for the term: *old*.

EXERCISE 4.3

Assuming a variable-length code is adopted to build the index term, encode the following posting lists. A posting list comprises pairs of (document ID, term frequency)

- i. (2,1), (4,2), (5,2) -> unary encoded)
- ii. (6,3), (7,1), (10, 4) -> Elias' gamma encoded)

EXERCISE 4.4

Assuming a variable-length code is adopted to build the index term, decode the following sequences. Each sequence represents a posting list for a given term, and comprises pairs of (document ID, term frequency)

- i. 0101001110111101111100 (unary encoded)
- ii. 1000110001101011001100 (Elias' gamma encoded)