

Homework Assignment 8

Due date: 20th of January 2014

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 10: INTRODUCTION TO WEB RETRIEVAL

EXERCISE 10.1

What are the major differences between web retrieval and classical IR?

EXERCISE 10.2

How can you estimate the size of a search engine's maintained index?

EXERCISE 10.3

What are the major problems WebCrawlers face?

EXERCISE 10.4

How can you estimate the size of the World Wide Web? Would it reflect the real size of the web and why?