

Homework Assignment 5

Due date: 17th of December 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 6: LANGUAGE MODELS

Document 1	The house in the town had the big old keep
Document 2	In the big old house in the big old gown
Document 3	The old night keeper keeps the keep in the town

Given the above collection of documents. Use MLE (Maximum likelihood estimator) to estimate the corresponding language model of Document 1 by:

EXERCISE 6.1

Using a unigram model coupled with simple smoothing, where $\alpha = 1$.

EXERCISE 6.2

Using a unigram model coupled with Linear smoothing, where $\lambda = 0.5$.

EXERCISE 6.3

Using a unigram model coupled with Ponte and Croft smoothing.

EXERCISE 6.4

Referring to the language model you estimated for document 1 in exercise 6.1, what's the probability that the corresponding language model generated the following query: *old house keep?*