

Information Retrieval and Text Mining: Assignment 8.

Problem 1.

Finish Assignment 7. Ask if you have any questions from previous assignments. [20 min]

Problem 2.

This exercise is about implementing Language Models. The corpus used is the CACM corpus you have seen earlier and which has relevance judgements for 63 queries. Go through the code carefully.

You will implement the models with

1. Jelinek-Mercer smoothing.
2. Dirichlet smoothing.

Print the results and evaluate it using Galago. Compare the evaluation scores for different parameter values with evaluation scores obtained with Galago.

Exercise date: Friday, January 20, 2012, 14:00.