Information Retrieval and Web Search Engines
Summer Semester 2010

Prof. Dr. Wolf-Tilo Balke and Joachim Selke

ifis
Institut für Informationssysteme
Technische Universität Braunschweig

Homework Assignment 4

Due to June 17, 2010
(35 points in total)

Remember: (1) Start early to work on this assignment. (2) Let us know if you need help.

Note: Again, for all of the following exercises, please use the *stemmed* version of the Reuters collection, which is available for download on the lecture website.

Exercise 4.1 (Retrieval with Language Models)

Implement a retrieval system based on the simple unigram language model presented in the lecture. Use linear smoothing with $\lambda = 0.8$. Use your system to answer the query "taxes reagan" (don't forget stemming). What is your subjective opinion about the result quality?   (5 points)

Exercise 4.2 (Relevance Judgments: Creating Test Data)

As we have seen in the lecture, an IR system's effectiveness is typically evaluated against some human-defined *ground truth,* that is, a test data set consisting of (1) a document collection, (2) a set of queries, and (3) a relevance judgment for each query–document pair.

Use the pooling method (based on the retrieval methods we used so far in the homework, namely, Boolean retrieval, vector space retrieval with TF-IDF and cosine similarity, 100-dimensional LSI with log entropy and cosine similarity, Binary Independence Retrieval with the 0.9 heuristic, and language models as used in the previous exercise) to determine which documents are relevant with respect to the following queries:

  a) taxes reagan

  b) oil price collapse

  c) toxic cargo

For each of these queries, please explain (1) which (more detailed) information need you assume to underlie the query and (2) what criteria your relevance judgments are based upon.

Hint: To simplify this task, we recommend to implement some helper functions in MATLAB: First, for each retrieval method, a function that takes a query as input and returns a ranking of the collection's documents along with the corresponding scores. Second, a function that takes a query as input, calls all retrieval methods, processes their result lists (by considering only a reasonably long prefix of each list), and create a duplicate-free list of documents, which have to be evaluated manually for relevance.   (15 points)

Exercise 4.3 (Evaluating Precision and Recall)

Use the test data set created in the previous exercise to evaluate the effectiveness of the five retrieval methods we used so far (see above). For each of the three queries, draw a picture showing the precision–recall at k curves for all five methods (you don't need to apply interpolation).

Compare the three pictures and discuss the strengths and weaknesses of each retrieval method.

Hint: For Boolean retrieval, drawing a precision–recall at k curve might be difficult. What can you do instead?

Hint 2: Drawing precision–recall at k curves is problematic if many documents get assigned the same score in a ranking. What can be done to avoid this problem?

(15 points)