CSC 849: Information Retrieval
Assignment 4
Relevance Feedback, and Retrieval Evaluation
100 Points

This assignment will provide you experience with pseudo relevance feedback, and the classic IR evaluation methodology. It consists of four parts.

1. Relevance Evaluation

For this step the task is to evaluate each of the 10 documents in `documents.txt` for its relevance to each of the three queries in file `queries.xml` We will use simple binary relevance scale: a document is either *relevant* or *non-relevant* to the query.
Write your relevance judgments to a file named:`[yourLastName]-qrels.txt`, and use the following format for each line of this file:
`[QryID] 0 [DocID] [Relevance]`

1. first element is numeric query ID,
2. second element is the number 0 for every line,
3. third is the numeric document ID,
4. fourth is your relevance judgment for the queryID, docID pair. 1 is relevant and 0 is non-relevant,
5. each element is separated by a single space character.

The total number of lines in this file should be 30 (10 per query).

Complete this task by Monday, Oct 17, and post your relevance judgment to the ilearn forum named "HW4: Relevance Judgments".
Inter-annotator agreement: Download the relevance judgment files of your fellow classmates from the ilearn forum, and write a script/program to compute the kappa statistic between your file and every other file.

2. Pseudo-relevance feedback (PRF)

Extend the tf.idf based scoring function that you implemented in homework 2 as follows:

1. The document retrieved at rank 1 (d_r) is assumed to be relevant to the query.
2. Compute the tf.idf score for each of the unique terms in d_r, and select the top x terms for query expansion.

3. Rerun the tf.idf based scoring function now with the {original query + x additional terms}, and output the new search results with the following format for each line:
   `[QryID] 0 [DocID] [Rank] [Score] tfidf`

3. Evaluation

Here we will bring together the previous two parts.

1. Experiment with a range of values for the parameter x, specifically, x = {1, 3, 5}.
2. For a given parameter setting x, concatenate the result files for the 3 queries. This should produce 3 result files, one for each of the parameter settings.
3. Evaluate the quality of each result file using the `trec_eval` program, available at http://trec.nist.gov/trec_eval/

   You will need your relevance judgments file to evaluate the quality of the search results using trec_eval.

4. Report

Write a report that:

1. Tabulates the search results' quality in terms of P@5, P@20, Recall@5, Recall@20, MRR (recip_rank) and MAP for each of the 3 parameter settings runs.
2. Summarizes your observations about PRF and its parameter x.
3. Reports all the kappa statistic scores that you computed in Part 1.

Upload to ilearn the following files:

1. report, (55 points)
2. code for computing kappa statistic, (15 points)
3. code for PRF (30 points)