CS6200 Information Retrieval

Homework5: Relevance Assessments, IR Evaluation

Objective

In this assignment, you will continue the work from previous HW by evaluating your vertical search engine. You will continue to work within the team you formed earlier.

You will be given queries for your topical crawl. Manual relevance assessments have to be collected, using your vertical search engine and a web interface.

You will have to code up the IR evaluation measures, essentially rewriting trec*eval. It is ok to look at the provided trec*eval code, but you have to write your own.

Assessments and IR Evaluation.

Obtaining queries

Each team will be assigned 3-4 queries specific to the topic you worked on HW3. The queries are going to show up in your Dropbox team file, or you can obtain them from the TAs.

Assessment graphical interface

In order to assess relevance of documents, you will have to create a web interface that displays the topic/query, and a given document list. Each URL in the list should be clickable to lead to the document text, pooled either from ES raw-html field or live from original URL. You would probably take as a start the web GUI you used for vertical search in HW3.

The interface has to contain an input fields for each URL/snippet in order for the assessor to input a 3-scale grade “non-relevant”, “relevant”, “very relevant” (or 0,1,2). The can be an input checkboxes, radio boxes, dropdown list, text input box, etc. The interface should also record the assessor ID (by name). You can add a “submit” button somewhere, and a count of how many documents have been assessed.

The input assessments should be stored in a QREL file (txt format) as

QueryID AssessorID DocID Grade

QueryID AssessorID DocID Grade

…..

where DocID is the canonical URL, AssessorID is your name (no spaces, like "Michael_Jordan"), Grade is one of {0,1,2}. You can temporarily store information in ES or a database if that is easier for you.

While not ideal, we are aware of some students are not being versed in Web-Development. So we will allow for the input to be manual directly to the QREL file. Thats is, you will use the vertical search web interface from HW3 with no input fields, and manually copy-paste the IDs into the qrel together with the assigned relevance grade.

You will have to demo your assessment process (the interface and the recording of grades).

Manual assessments.

Each student has to manually assess about 200 documents for each query. So if your team has 3 queries, each student will assess 600 documents, and each document-per-query will be assessed three times (assuming three team members).

The QREL file should record all the assessments and be placed in your Dropbox folder when you are done.

Write your own trec_eval

Write a program that replicates trec_eval functionality. Input : a ranked list file and QREL file, both in TREC format. Both files can contain results and assessments for multiple queryIDs.

First, sort docIDS per queryID by score. Then for each query compute R-precision, Average Precision, nDCG, precision@k and recall@k and F1@k (k=5,10, 20, 50, 100). Average these numbers over the queryIDs. If run with -q option your program should display the measures for each queryID before displaying the averages.

Run your trec*eval on HW1 runs with the qrel provided to confirm that it gives the same values as the provided trec*eval.

Run your trec_eval on the HW3 vertical search engine.

Precision-Recall Curves

For each one of the HW4 queries, create a precision-recall plot.

Extra Credit

These extra problems are provided for students who wish to dig deeper into this project. Extra credit is meant to be significantly harder and more open-ended than the standard problems. We strongly recommend completing all of the above before attempting any of these problems.

Points will be awarded based on the difficulty of the solution you attempt and how far you get. You will receive no credit unless your solution is “at least half right,” as determined by the graders.

EC1:

A nice, fast, web interface for evaluation, will get you EC points.

Also EC points will be given for highlighting query terms in the document source when the entire document is displayed.

Rubric

15 points
A proper interface for assessment, including keeping track of assessments
30 points
About 600 manual assessments, for 3 or 4 queries on your topic (qrel file in Dropbox)
30 points
Your trec_eval
25 points
Precision Recall curves