

Dublin City University
School of Computing
CA4009: Search Technologies
Laboratory Session 2

November 2016

Module Coordinator: Gareth Jones

Laboratory Tutors: Piyush Arora, Wei Li

1 Introduction

This laboratory introduces working standard information retrieval test collections which enable the effectiveness of information retrieval systems and algorithms to be evaluated using standard metrics. The design and construction of information retrieval test collections is introduced in Section 3: Text Retrieval of the lectures notes from CA4009 Search Technologies.

You should prepare a report similar to the one from Laboratory 1 and submit it via loop. Similar to Laboratory 1, you can submit your report at any time until before the start of the next laboratory at 10.00 next Thursday.

2 Working with the TREC Topic Set

2.1 Manual examination of TREC data and search results

For the first part of the second laboratory you will work with a TREC topic set and the corresponding qrel relevance information, together with a sample TREC formatted “run” . The ‘run’ contains a list of the top 1000 ranked documents for each query created using a combination of the Title and Description fields for each topic using “language modelling” (we didn’t cover this method in the module) as the ranking function.

A single archive file containing all these files can be downloaded from: <http://computing.dcu.ie/~dganguly/trec.tar.gz>.

- Download and extract the files from the archive file. The topics file: topics.301-450.xml. The sample output file: trec678.res. The relevance assessments: qrels.trec678.adhoc.
- Briefly examine the format and contents of the files. You should be able to see the structure of the standard TREC topics search queries) and their individual IDs (document numbers); locate the query IDs, the document IDs (unique numbers) and matching scores in the sample run file; the document IDs in the qrel file and the indication of whether the document is relevant or not (the relevance indicator is set to 0 or 1).

You might find it help to consult the following URL for a more formal explanation of what you can see in these files: http://faculty.washington.edu/levow/courses/ling573_

SPR2011/hw/trec_eval_desc.htm. Some further information from stackoverflow is available from here: <http://stackoverflow.com/questions/4275825/how-to-evaluate-a-search-retrieval-engine-using-trec-eval>.

- Copy and paste some examples of Title, Description and Narrative fields into the query box in the search interface. Examine the list of retrieved results in each case looking for documents identified as relevant in the qrel. Look for variations in the rank positions of these relevant documents for queries created using the different fields from the TREC topic. Can you provide any explanation for these different rank positions based on the contents of the queries. Enter your observations in a new report document for laboratory 2. You should examine at least 3 different topics.

3 Exploring Evaluation Metrics

For the experiments in this section of the laboratory you will be using the sample Lucene output file downloaded in the previous section “trec678.res”.

The trec_eval software itself must be downloaded from here: http://trec.nist.gov/trec_eval/.

To use the trec_eval software you need to download the archive file, and run the makefile to build a local binary file in your Linux file space.

- Download and build trec_eval in your linux file space. You can run trec_eval directly in the directory where you compile or, if you prefer, install it as a binary in a bin directory (if you don't know how to do this, don't worry, just use your compiled copy in your current directory).
- Run trec_eval using the sample files provided in the “test” folder with the download file for trec_eval using the following command

```
trec_eval qrels.test results.test
```

You should observe the standard trec_eval output showing the various evaluation metric values calculated by trec_eval for the sample run data with the corresponding qrel data.

- Now run trec_eval for the sample retrieval run file “trec678.res” provided for the laboratory and the corresponding qrel file using the following command.

```
trec_eval qrels.trec678.adhoc trec678.res
```

This will produce a set of standard information retrieval metric results for the TREC document collection using the language modeling ranking function for this set of topics with the manually judged relevance data. trec_eval produces results for a wide range of metrics, You should concentrate on the ones giving details of the number of documents retrieved, number of relevant documents retrieved, precision at rank cut offs (Prec), mean average precision (MAP) and recall.

If you need to, consult your lecture notes for details of how these are calculated.

The results here are averaged over the whole topic set.

Examine the values of Precision at rank cutoffs, Mean Average Precision (MAP) and Recall, and comment on your observations in your laboratory report.

- Re-run trec_eval using the following command:

```
trec_eval -q qrels.trec678.adhoc trec678.res
```

You should find that adding the `-q` flag here generates a much larger number of results. What this shows you is the results for each individual topic as well as the averaged results.

Examine the results for some of the individual queries from the TREC sample file, and comment on your observations in your laboratory report. You should look for details such as the Precision values and the number of relevant documents available and retrieved for each query. Using the definitions of precision and recall, you should be able to comment on the effectiveness of retrieval for individual topics at different ranks for the different topics.

4 Exploring Relevance Assessment

- Select 3 topics.
- For each of these topics copy the Title field of the query into the search box in the interface one by one and run search using one of the ranking functions.
- Obtain the ranked output for the current query.
- Open each of the top 10 ranked documents for the query.
- Read all three fields of the topic, and determine whether you would judge the document to be relevant to the topic. Record your judgements in your laboratory report.
- Compare your judgements with those in the `qrel` file for this topic. Do your judgements of relevance for these topics correspond with those in the `qrel` file. Documents not present in the `qrel` file for this topic, have not previous been judged for relevance, and can be regarded as judged to be non-relevant by the TREC organisers. Comment on your findings in your laboratory report.