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

November 2019

Module Coordinator: Gareth Jones

Laboratory Tutors: Tianbo Ji, Abhishek Kaushik, Yasufumi Moriya, Procheta Sen

1 Introduction

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

2 Laboratory Report and Submission

Similar to Laboratory 1, you should create a electronic report file for this laboratory.

- Include the title and date of the laboratory, and your names at the beginning of your report.
- For each activity described below, enter your answers into your report file, making clear which section your response relates to.
- At the end of the laboratory session you should upload your report to the CA4009 loop page via the link for this laboratory.
- If you do not finish the exercises to your satisfaction, you can complete the assignments in your own time, and submit an extended report. The latest date for submission of the extended report is one week after the laboratory session.

3 Manual examination of a TREC test collection and sample 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 output of a search "run" . The 'run" contains a ranked list of the top 1000 scoring documents for each query created using a combination of the Title and Description fields for each topic created using a "language modelling" information retrieval methods. (we didn't cover this method in the module) as the ranking function.

A single archive file containing all the files that you need for this exercise can be downloaded from:

`http://computing.dcu.ie/~dganguly/trec.tar.gz`

What you need to do:

- Download and extract the files from the archive file.
- Uncompress this archive file. You can do this using the tools provided provided in the linux interface or using the commands:

  `gunzip trec.tar.gz`

  `tar xvf trec.tar`

  This should create a folder containing the following files:

  - Topics file: topics.301-450.xml. 150 TREC format topic statements
  - Sample output file: trec678.res. Sample output for the 150 TREC topics in standard TREC output format.
  - Relevance assessments: qrels.trec678.adhoc. Manual relevance assessments for the 150 TREC topics.

- Briefly examine the format and contents of the files.

  In the topics file you should be able to see the structure of the standard TREC topics (search queries) and their individual IDs (topic/query numbers) as described in the notes for Labotatory 1.

  Locate the query IDs, the document IDs (unique numbers) and matching scores in the sample run file.

  You should be able to find the document IDs associated with each of the topics 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 helpful 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` which you used in the first laboratory.

  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 available at: URL `http://136.206.48.37:8084/IRModelGenerator/search.jsp`.

  This query interface enables you to search the same document collection as used to create the ranked output lists in the example that you downloaded and the qrel file for these set of search topics that you downloaded.

  Examine the list of retrieved results in each case looking for documents identified as relevant in the qrel file for this topic. 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 words contained in the queries created using the different fields and and the contents of the retrieved documents?

  You should examine at least 3 different topics.

4 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" with the standard information retrieval evaluation application `trec_eval`.

The trec_eval software itself must be downloaded from here:

`http://trec.nist.gov/trec_eval/`

There are two versions of trec_eval available at this URL. You should download and install the

`trec_eval_latest.tar.gz` version.

What you need to do:

- Download the `trec_eval` file above.
- Extract the contents of the download file as before.
- Compile the `trec_eval` application n your Linux file space to make an operational program.

  To do this, once you have extracted the folder, you need to navigate to the folder, and run the command

  `make trec_eval`
- You can run `trec_eval` directly in the directory where you compile it using the command

  `./trec_eval`

  or, if you prefer and you know how to do this, you can install is as a binary in a bin directory, and add the directory where you locate it to your linux `$PATH`.
- Run `trec_eval` using the sample files provided in the "test" folder provided with the downloaded `trec_eval` program using the following command

  `trec_eval qrels.test results.test`

  This should produce the standard `trec_eval` output showing the various evaluation metric values calculated by `trec_eval` for the sample run data in the files `results.test` with the corresponding qrel data in `qrels.test`.
- 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`

  To do this, copy the compiled `tec_eval` application to the folder containing "trec678.res" and "qrels.trec678.adhoc".

  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 metrics 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 and what you understand the results to mean, in your report for this laboratory.

- 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 method that was used to create these results for individual topics at different ranks.

5 Exploring Consistency of Relevance Assessment for Topics

- Select 3 topics from the topic set.
- For each of these topics copy the Title field of the topic into the search box in the search interface one by one and run search for this query 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 each document that you open for the query to be relevant to the topic. Record your judgements in your laboratory report. with the document IDs.
- Compare your relevance judgement for each document with those that you find in the qrel file for each of your chosen topics.

  Do your judgements of relevance for these topics correspond with those in the qrel file? Based on the details of the topic statement and the document can you see why you disagree with the original assessment of relevance in the qrel file?

  Documents not present in the qrel file for each topic were not judged for relevance when the qrel file was created for TREC, and can be regarded as judged to be non-relevant by the TREC organisers.

  Give details of your investigation in your laboratory report, include details of topics used, ranking function selected, documents IDs opened, agreement and non-agreement of relevance for each document for each query.