

Offline Evaluation Hands-on

1 Introduction

In this hands-on you will use offline evaluation to assess the quality of different search methods. You will use the following data and tools:

- TREC test collection from the first two years of TREC (TREC 1-2).¹ This collection has the following components:
 - Test documents: mainly journal articles from Associated Press (1988-1989), Wall Street Journal (1987-1992) and others
 - 50 test queries (called “topics” in TREC)
 - Binary relevance labels for these queries (called “qrels” in TREC)
- Terrier retrieval engine.² This tool provides the means for indexing and retrieving documents. We will discuss both notions later in the course.
- An official TREC evaluation tool `trec_eval`.³

2 Setup

Prepare the data and tools as follows:

1. Download the data and Terrier with an already created index from <https://yadi.sk/d/7xXt0GBQ3E4EyA>.
 - (a) Unpack the data and get acquainted with it. You will not use it directly in this hands-on, but will need it later.
 - (b) Unpack Terrier. You will work with it throughout this hands-on. The above data is already indexed for you in the `var/index` folder. You will also create different indices yourself later in the course.
2. Download `trec_eval` from http://trec.nist.gov/trec_eval/ and compile it by typing `make` in the corresponding folder.

3 Hands-on

Now you will run different search methods and evaluate their results using the offline evaluation paradigm.

1. Retrieval
 - (a) Get acquainted with the “Querying” part of the page http://terrier.org/docs/v4.2/quickstart_experiments.html. Do some querying in the interactive mode.

¹<http://trec.nist.gov>

²<http://terrier.org/>

³http://trec.nist.gov/trec_eval/

- (b) In the Terrier folder, run the following command to retrieve document for 50 test queries:

```
bin/trec_terrier.sh -r -Dtrec.topics=<path to the file with TREC topics>
-Dtrec.model=<search method>
```

Be careful, the underscore in trec_terrier.sh disappears when copying the command!

- (c) The list of available search methods is available at http://terrier.org/docs/v4.2/configure_retrieval.html under “Weighting Models and Parameters”. The results for each method are stored in `var/results/<search method>.res`.
- (d) Run different search methods.

2. Evaluation

- (a) Look through the README file of trec_eval.
- (b) Run the following command to evaluate the retrieval results of a search method:

```
trec_eval -q -m all_trec <path to the file with TREC qrels> <path to the file with search results>
```

Be careful, the underscore symbols disappear when copying the command!
- (c) The script outputs the evaluation results for each test query and the results averaged over 50 queries.
- (d) Evaluate different search methods and compare the results.