

NDAK15005U Information Retrieval (IR) 2019 - Assignment 2

**Deadline: 27 May, 23h55. Submissions
must be anonymous (no name, no KUid)**

IR course, 2019

May 20, 2019

This assignment has the following three learning objectives:

1. Learn to evaluate existing distributional semantics models (knowextend ad-hoc text retrieval models with word embeddings (33% of the whole assignment grade)).
2. Learn to evaluate ad-hoc text retrieval models (33% of the whole assignment grade).
3. Learn to present experimental results in a scientific report (34% of the whole assignment grade).

1 Word embeddings for ad-hoc text retrieval

In the previous assignment we considered how word embeddings can be applied to document classification and similarity tasks. This assignment extends the skills you acquired in assignment 1 to a more challenging task: apply word embeddings to ad-hoc text retrieval.

In this assignment you will use the robust04 collection (download from Absalon). The collection consists of 500,000 documents and 250 queries, where queries are labelled with query relevance scores (qrels) for a subset of the documents in the collection. The task of ad-hoc retrieval is to rank the documents decreasingly by their relevance to a given query. Since few queries with relevance scores are available for this task, work on this collection mostly focuses on tuning algorithms with few parameters. In this assignment you are going to do exactly this. For the programming part of the assignment you should follow these steps:

- Look closely at the collection, queries, relevance scores, how they are stored, and how a ranking should be evaluated. We strongly suggest to use `trec-eval`¹ for evaluation.
- Split the queries into train and test sets (80% and 20%). Split the train set into 5 folds used in cross-validation for parameter tuning.
- For evaluation purposes you need to report the Mean Average Precision (MAP) and average precision at cut-offs of {5,10,20}.
- Tune and run BM25. You are allowed to use existing search libraries².
- Word embeddings can be used to extend the query itself, without modifying the ranking function. Use word embeddings to do query expansion as done by Kuzi et al.³. You should implement both the centroid approach and at least one fusion approach. You can use BM25 as the ranking function.
- Word embeddings can also be integrated directly into the ranking function itself as done by Kenter and de Rijke⁴. You should implement the ranking function described in section 3.1.1 of that paper, but do not have to implement anything else from the paper.

When this is done you should have 3 distinct ranking functions, BM25, query extended BM25, and BM25 with word embeddings integrated. If you face problems due to limited computational resources⁵, you are allowed to use BM25 to do an initial ranking, and then rerank the top 10000 documents with the word embedding based approaches. If you do this you should discuss the potential consequences compared to no initial ranking.

1.1 Written report

You should structure your report like a scientific research article. You can seek inspiration from the papers referenced for the two embedding based methods on how to do this. You need to have at least the following sections:

- Abstract: Summarize your work and findings.
- Introduction: Describe the problem of ad-hoc retrieval and explain the purpose of using word embeddings for this task.
- Method: Describe the 3 ranking functions (including the mathematical formulations) you have implemented.

¹ https://github.com/usnistgov/trec_eval

² We suggest to use the following <https://github.com/castorini/anserini/blob/master/docs/experiments-robust04.md>

³ http://publish.illinois.edu/saar-kuzi/files/2017/10/w2v_cikm16.pdf

⁴ <https://staff.fnwi.uva.nl/m.derijke/wp-content/papercite-data/pdf/kenter-short-2015.pdf>

⁵ Consider using the free <https://colab.research.google.com/>

- Experiments: Describe how you have performed the experiments, such that other people are able to reproduce them.
- Results: Present the results.
- Discussion: Discuss the strengths and weaknesses of the 3 approaches.