Evidence Combination

This module combines the output from several matchers into a single retrieval result and subsequently it applies the cutoff to the combined results.

Input

The input is the output from one or several matchers for a single query in trec format, a processed query, the configuration file and a path to the data directory.

Output

Combined outputs of the matchers for a single query in a tsv format (with the cutoff applied).

Docker Commands

```
docker run \
    -e INDIRS=dir1+dir2 \
    -e QUERY=query100 \
    -e CONFIG=config_file.json \
    -v experiment_directory:/media/input/experiments \
    -v directory_with_queries:/media/input/queries \
    -v directory_with_config_file:/media/input/configs \
    -v root_data_directory:/media/input/data \
    -v output_directory:/media/output \
    ...
```

- `INDIRS`: names of the directories with the outputs of the matchers with trec files. Multiple directories are divided by '+'
- `QUERY`: filename of the pre-processed query from the output of the query analyzer
- `CONFIG:` filename of the experiment configuration file
- `Experiment_directory:` root directory of the experiment, which contains the INDIRS
- `Directory_with_queries:` Directory which contains the processed query
- `Directory_with_config_file:` Directory which contains the configuration file
- `Root_data_directory:` Root directory path with the data
- `Output_directory:` output directory filename

Examples

docker run \
-e "INDIRS=UMD-CLIR-workQMDir-EdiNMT+UMD-CLIR-workQMDir-umdSMT" \
-e "QUERY=query100" \
-e "CONFIG=swahili_config_new_pipeline_new_langid.json" \
-v
/storage2/proj/badrashiny/CLIR/matcher5-out/query-analyzer-umd-v4.0_matching-umd-v5.0_evi

dence-combination-v2.0/TL_QUERY1_ANALYSIS1_umdSMT_PSQ_EdiNMT_Cutoff60:/media/input/experiments \
-v
/storage2/proj/badrashiny/CLIR/matcher5-out/query-analyzer-umd-v4.0_matching-umd-v5.0_evidence-combination-v2.0/TL_QUERY1_ANALYSIS1_umdSMT_PSQ_EdiNMT_Cutoff60/UMD-CLIR-workQPDir:/media/input/queries \
-v /storage/proj/petra/configs/tests/:/media/input/configs \
-v /storage2/data/NIST-data/:/media/input/data \
-v /storage/proj/petra/experiments/combination:/media/output \
--name evidence-combination-instance evidence-combination /bin/bash

System Requirements

- CPU
- RAM
- GPU
- GPU-RAM
- Target CUDA version and/or minimum NVIDIA driver version (current Scripts servers have CUDA: 11.2 and Driver: 460.32.03)

Standalone

yes

Approach

The component supports several different combination types: weighted CombMNZ combination, reciprocal-rank-based combination and Borda count combinations. In the case of the CombMNZ combination, the scores might be first normalized using Sum-to-one (STO) normalization.
The component also supports different cutoff approaches: STO cutoff (with a cutoff threshold parameter), query specific threshold and ranked threshold. It is also possible to use an average of these two or three approaches and a maximal cutoff limit which might be applied when the average is used.