TREC CAR 2017 Run Summary

CUISER

Team	Task	Run type	Ranking methods
CUIS	entity	automatic	BM25 Sequential Dependence Model

Features

- Trained on `train-v1.5`

Description

Replace `para_id` in passage ranking results by `article_id`

CUISPR

Team	Task	Run type	Ranking methods
CUIS	passage	automatic	BM25 Sequential Dependence Model

Features

- Trained on `train-v1.5`

Description

Scoring pass

1. BM25 by Lucene 6 scoring pass
2. a slightly modified sequential dependence model

ECNU-runONE

Team	Task	Run type	Ranking methods
ECNU	passage	automatic	BM25

Features

- Trained on `train-v1.5`
- Uses learning-to-rank.

Description

First, we use lucene to select candidate paragraphs, and then we use bm25 score and word matching as features to training a ranking model with Ranklib.

treccarict

Team	Task	Run type	Ranking methods
ICTNET	passage	automatic	BM25

Features

- Uses a dump of Wikipedia
- Unsupervised

Description

Use BM25 for retrieval.

mpii-nn4_pos_hperc

Team	Task	Run type	Ranking methods
MPIID5	passage	automatic	BM25

Features

- Uses data provided as `unprocessedtrain` or `halfwiki`
- Uses pre-trained word embeddings
- Is trained with data provided as `train-v1.5`

Description

PACRR variant on BM25 candidate list.

mpii-nn6__pos

Team	Task	Run type	Ranking methods
MPIID5	passage	automatic	BM25

Features

- Uses data provided as `unprocessedtrain` or `halfwiki`
- Uses pre-trained word embeddings
- Is trained with data provided as `train-v1.5`

Description

PACRR variant on BM25 candidate list.

mpii-nn6__pos__tprob

Team	Task	Run type	Ranking methods
MPIID5	passage	automatic	BM25

Features

- Uses data provided as `unprocessedtrain` or `halfwiki`
- Uses pre-trained word embeddings

- Is trained with data provided as `train-v1.5`

Description

PACRR variant on BM25 candidate list.

nyudl-ds

Team	Task	Run type	Ranking methods
NYUDL	passage	automatic	BM25

Features

- Uses data provided as `unprocessedtrain` or `halfwiki`
- Uses pre-trained word embeddings
- Uses neural network technology
- Uses learning-to-rank
- Is trained with data provided as `train-v1.5`

Description

Simple Document Classifier using avg word embeddings in the documents as document vector and last hidden state of an LSTM as query vector. 2-layer feed forward neural net to select which documents are relevant given a query.

nyudl-qr

Team	Task	Run type	Ranking methods
NYUDL	passage	automatic	BM25

Features

- Uses a dump of Wikipedia
- Uses pre-trained word embeddings

- Uses neural network technology
- Uses learning-to-rank
- Uses other supervised machine learning method
- Is trained with data provided as `train-v1.5`

Description

Query reformulation with deep reinforcement learning.

nyudl-qrds

Team	Task	Run type	Ranking methods
NYUDL	passage	automatic	BM25 Neural network classifier

Features

- Uses a dump of Wikipedia
- Uses pre-trained word embeddings
- Uses neural network technology
- Uses learning-to-rank
- Is trained with data provided as `train-v1.5`

Description

Query reformulation using reinforcement learning + Lucene + Neural Net Classifier to select documents

UNH-benchmarkY1test.bm25

Team	Task	Run type	Ranking methods
TREMA-UNH	passage	automatic	BM25

Features

- Unsupervised

Description

BM25 run using the concatenation of heading, parent headings and page title as keyword query

UNH-benchmarkY1test.expan

Team	Task	Run type	Ranking methods
TREMA-UNH	passage	automatic	BM25 Query expansion (entities and section headings)

Features

- Uses a linguistic graph such as Wordnet
- Uses entity linking
- Uses data provided as ’unprocessedtrain’ or ’halfwiki’
- Uses other supervised machine learning method
- Is trained with data provided as `train-v1.5`
- Is trained with data provided as `test200`

Description

BM25 run using the concatenation of heading, parent headings and page title as keyword query. In addition query expansion with two sources: 1. the query is entity linked with TagMe. Terms from the first paragraph of the entity are used for expansion (like RM3) 2. if the same heading is contained in another article, then expansion terms from these sections will be used for expansion (like RM3).

This method is in spirit of the WikiKreator system. Balancing parameters are manually adjusted on `test200`

top100-c-pr-bm25

Team	Task	Run type	Ranking methods
TREMA-UNH	entity	automatic	BM25 PageRank

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.
- Uses data provided as 'unprocessedtrain' or 'halfwiki'
- Unsupervised
- Is trained with data provided as `benchmarkY1train`

Description

Ranks entities by degree centrality on a sub-graph that is extracted as follows: 1. edges in KG are associated with paragraph-long text 2. A BM25 model is used to retrieve edges in response to the query 3. edges are weighted according to their frequency 4. PageRank on this weighted graph is used to rank entities Support paragraphs are taken from the the paragraph associated with the entity's highest ranking edge.

top100-rr-marg-bm25

Team	Task	Run type	Ranking methods
TREMA-UNH	entity	automatic	BM25 Degree centrality

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.
- Uses entity linking
- Uses data provided as `unprocessedtrain` or `halfwiki`
- Does not make use of any training data (unsupervised)
- Is trained with data provided as `train-v1.5`

- Is trained with data provided as `benchmarkY1train`

Description

Ranks entities by degree centrality on a sub-graph that is extracted as follows: 1. edges in KG are associated with paragraph-long text 2. A BM25 model is used to retrieve edges in response to the query 3. edges are weighted according to their reciprocal rank 4. DegreeCentrality on this weighted graph is used to rank entities Support paragraphs are taken from the the paragraph associated with the entity's highest ranking edge.

top100-sc-ppr-bm25

Team	Task	Run type	Ranking methods
TREMA-UNH	entity	automatic	BM25 Personalized PageRank

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.
- Uses entity linking
- Uses data provided as 'unprocessedtrain' or 'halfwiki'
- Does not make use of any training data (unsupervised)
- Is trained with data provided as train-v1.5
- Is trained with data provided as benchmarkY1train

Description

Ranks entities by degree centrality on a sub-graph that is extracted as follows: 1. edges in KG are associated with paragraph-long text 2. A BM25 model is used to retrieve edges in response to the query 3. edges are weighted according to their frequency 4. seed nodes are retrieved from an entity index of `unprocessedtrain` using BM25 5. PersonalizedPageRank on this weighted graph with seed nodes is used to rank entities Support paragraphs are taken from the the paragraph associated with the entity's highest ranking edge.

UTDHLTRIAR

Team	Task	Run type	Ranking methods
UTDHLTRI	passage	automatic	BM25 TF-IDF Dirichlet Language Model

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.
- Uses entity linking
- Uses learning-to-rank
- Is trained with data provided as test200
- Is trained with data provided as benchmarkY1train

Description

1. Initial Search
    1. Parse the article title and section names to use in part of query for Lucuene search using BM25
    2. Use entities extracted from Dbpedia long abstracts data set from Dbpedia spotlight tagging to augment query
2. Re-Ranking
    1. use Ranklib's Ada Rank implementation to re-rank paragraphs

UTDHLTRINN20

Team	Task	Run type	Ranking methods
UTDHLTRI	passage	automatic	BM25 TF-IDF Dirichlet Language Model

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.

- Uses entity linking
- Uses pre-trained word embeddings
- Uses neural network technology
- Uses learning-to-rank
- Is trained with data provided as train-v1.5

Description

1. Initial Search
    1. Parse the article title and section names to use in part of query for Lucuene search using BM25
    2. Use entities extracted from Dbpedia long abstracts data set from Dbpedia spotlight tagging to augment query
2. Re-Ranking
    1. use neural learning to rank model to re-rank paragraphs from Initial Search

UTDHLTRINN50

Team	Task	Run type	Ranking methods
UTDHLTRI	passage	automatic	BM25 TF-IDF Dirichlet Language Model

Features

- Uses a knowledge graph such as DBpedia, Freebase, etc.
- Uses entity linking
- Uses pre-trained word embeddings
- Uses neural network technology
- Uses learning-to-rank
- Is trained with data provided as `train-v1.5`

Description

1. Initial Search

   1. Parse the article title and secition names to use in part of query for Lucuene search using BM25
   2. Use entities extracted from Dbpedia long abstracts data set from Dbpedia spotlight tagging to augment query
2. Re-Ranking
   1. use neural learning to rank model to re-rank paragraphs from Initial Search