

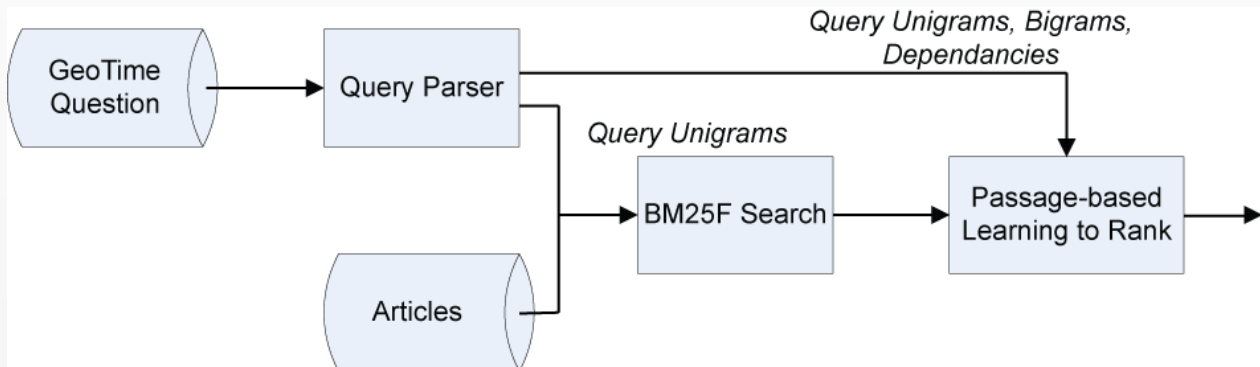
GeoTime Retrieval through Passage-based Learning to Rank

Xiaolin Wang, Hai Zhao, Baoliang Lu

BCMI Lab, Shanghai Jiao Tong University

arthur.xl.wang@gmail.com {zhaohai; blu}@cs.sjtu.edu.cn

The Brain-like Computing and Machine Intelligence Lab (BCMI) of Shanghai Jiao Tong University takes part in the NTCIR-9 GeoTime English subtask. A Passage-Based Learning to Rank (PGLR) method to approach this task.



Process GeoTime topics (find clues for GeoTime answers)

- Unigrams of DESCRIP. (*UoD*) – unigrams not in a customized stop word list.
- Bigrams of DESCRIP. (*BoD*) – pairs of *UoDs* which are neighbors with stop words skipped.
- Dependencies of DESCRIP. (*DoD*) – dependencies with (at least one unigram is *UoD*).
- Unigrams of NARRA. (*UoN*) – unigrams in *BoN* and *DoN*.
- Bigrams of NARRA. (*BoN*) – pairs of neighboring non-stop unigrams (at least one unigram is *UoD*).
- Dependencies of NARRA. (*DoN*) – dependencies with (at least one unigram is *UoD*).

Topic	Unigrams	Bigrams	Dependencies
(D)Where and when did the space shuttle Columbia disaster take place?	space shuttle columbia disaster	space-shuttle shuttle-columbia columbia-disaster	shuttle-space disaster-columbia shuttle-disaster
(N)The user wants to know when and in what state the space shuttle Columbia exploded.	state space shuttle columbia explode	state-space space-shuttle shuttle-columbia columbia-explode	columbia-space columbia-shuttle explode-columbia

Nine Features for passage-based learning to rank (unit: each paragraph)

- the document's BM25F score
- TFIDF similarities of DESCRIPTION's unigrams(*UoD*), bigrams(*BoD*), dependencies(*DoD*)
- TFIDF similarities of NARRATIVE's unigrams(*UoN*), bigrams (*BoN*), dependencies(*DoN*)
- whether contains temporary named entities
- whether contains geographic named entities

Runs	Doc Retrieval	NARRA.?	PGLR?	MAP	Q	nDCG@10
SJTUBCM-EN-EN-01-DN	BM25F	Yes	No	0.3326	0.3557	0.4498
SJTUBCM-EN-EN-02-DN	BM25F	Yes	Yes	0.3648	0.3884	0.5127
SJTUBCM-EN-EN-03-D	BM25F	No	No	0.2885	0.3111	0.4217
SJTUBCM-EN-EN-04-D	BM25F	No	Yes	0.3141	0.3370	0.4429