
Advanced Natural Language Processing and Information Retrieval

Answer sentence reranking pipeline

Alessandro Moschitti, Kateryna Tymoshenko

Email: moschitti@disi.unitn.it, k.tymoshenko@gmail.com

Overview

■ Software

- you may find the installation instructions and instructions how to run in the ANLPIR-2018 github
- The distribution contains the scripts/
experiment_launchers/experiment_launcher.py
script which automatically generates the commands to run and end-to-end experiment
- We will experiment on WikiQA



Description paper

- Tymoshenko, Kateryna, et al. "RelTextRank: An Open Source Framework for Building Relational Syntactic-Semantic Text Pair Representations." Proceedings of ACL 2017, System Demonstrations (2017): 79-84.



WikiQA corpus

- An answer selection corpus by Microsoft
 - Yang, Y., Yih, W.-T., & Meek, C. (2015). WIKIQA: A Challenge Dataset for Open-Domain Question Answering. Proceedings of EMNLP 2015, 2013–2018. Retrieved from <http://www.aclweb.org/anthology/D15-1237>
- **Corpus created manually:**
 - questions are Bing query logs
 - for each question an relevant Wikipedia page is manually retrieved, and sentences from its first paragram are used as answers



WikiQA corpus statistics

- The corpus has many question without a correct answer passage (only wrong ones) and vice versa
- The are irrelevant for evaluation (you select a correct passage when it is not there)
- Thus, there are three modalities:
 - RAW: keep everything (but since most questions are not answerable MAP will be < 40)
 - NO ALL NEG: remove questions which have no correct answers
 - NO ALL SAME: further remove questions which have only correct answers (there are < 10 of them)

	CORPUS SIZE					
	TRAIN		DEV		TEST	
	Q	A	Q	A	Q	A
RAW	2118	20360	296	2733	633	6165
NO ALL NEG	873	8672	126	1130	243	2351
NO ALL SAME	857	8651	121	1126	237	2341



qid	question	aid	answer	label
RQ1	how are glacier caves formed?	RQ1-D1-0	A partly submerged glacier cave on Perito Moreno Glacier .	False
RQ1	how are glacier caves formed?	RQ1-D1-1	The ice facade is approximately 60 m high	False
RQ1	how are glacier caves formed?	RQ1-D1-2	Ice formations in the Titlis glacier cave	False
RQ1	how are glacier caves formed?	RQ1-D1-3	A glacier cave is a cave formed within the ice of a glacier .	True
RQ1	how are glacier caves formed?	RQ1-D1-4	Glacier caves are often called ice caves , but this term is properly used to describe bedrock caves that contain year-round ice.	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-0	This tablespoon has a capacity of about 15 mL.	True
RQ16	how much is 1 tablespoon of water	RQ16-D16-1	Measuring Spoons	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-2	In the US and parts of Canada, a tablespoon is the largest type of spoon used for eating from a bowl.	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-3	In the UK, Europe and most Commonwealth countries, a tablespoon is a type of large spoon usually used for serving.	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-4	In countries where a tablespoon is a serving spoon, the nearest equivalent to the US tablespoon is either the dessert spoon or the soup spoon .	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-5	A tablespoonful, nominally the capacity of one tablespoon, is commonly used as a measure of volume in cooking .	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-6	It is abbreviated as T, tb, tbs, tbsp, tblsp, or tblspn.	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-7	The capacity of ordinary tablespoons is not regulated by law and is subject to considerable variation.	False
RQ16	how much is 1 tablespoon of water	RQ16-D16-8	In the USA one tablespoon (measurement unit) is approximately 15 mL; the capacity of an actual tablespoon (dining utensil) ranges from 7 mL to 14 mL.	True
RQ16	how much is 1 tablespoon of water	RQ16-D16-9	In Australia one tablespoon (measurement unit) is 20 mL.	True

Input data format

QID Text

Q1 how are glacier caves formed?
Q2 How are the directions of the velocity and force vectors related in a circular motion
Q5 how did apollo creed die
Q6 how long is the term for federal judges
Q7 how a beretta model 21 pistols magazines works
Q9 how a vul works

data/wikiQA/WikiQA-train.questions.txt

QID AnsID IR_rank IR_score label text

Q1 Q1-D1-0 0 0.00000 false A partly submerged glacier cave on Perito Moreno Glacier .
Q1 Q1-D1-1 0 0.00000 false The ice facade is approximately 60 m high
Q1 Q1-D1-2 0 0.00000 false Ice formations in the Titlis glacier cave
Q1 Q1-D1-3 0 1.00000 true A glacier cave is a cave formed within the ice of a glacier .

data/wikiQA/WikiQA-train.tsv.resultset



Running the experiments

- Follow the instructions on the github to run experiments and evaluate the performance
- You can compare your results to the SoA systems:

	No All-SAME	
	MRR	MAP
LCLR (Yang et. al., 2015} impl. of (Yih et al. 2013)	60.86	59.93
HybridTK-NN (Tymoshenko et al., 2017)	74.08	72.19
IWAN-att (NNs) (Shen et al., 2017)	74.37	72.62
Compare-aggregate (NNs) (Wang et al, 2017)	74.83	73.68
Cmp-aggr-k-max (NNs) (Bian et al., 2017)	75.80	74.78
Cmp-aggr-listwise (NNs) (Bian et al., 2017)	75.29	73.96
HyperQA (NNs) (Tay et al., 2018)	72.01	70.47

Most of the papers where evaluated on the questions with only positive answers as well (No ALL-NEG).

Our evaluator discards them. In this table we calculate the SoA system performances in the ALL-SAME mode.



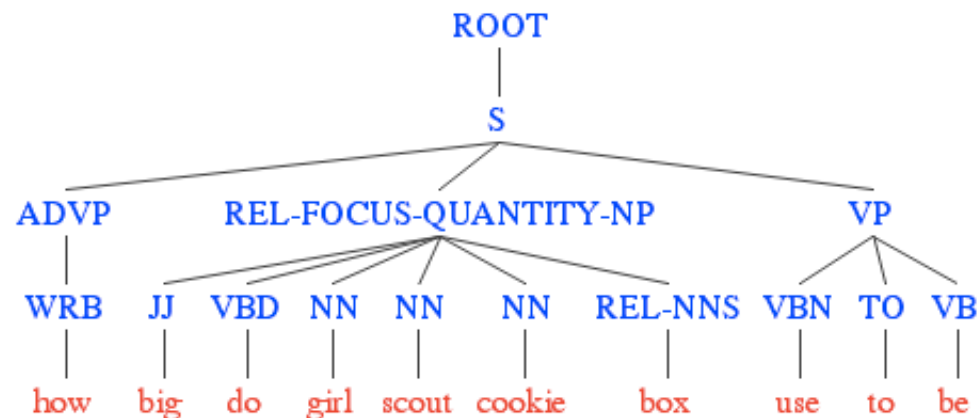
Structure vizualization

```
java -Xmx4G it.unitn.nlpir.system.demo.TextPairRepresentationDemo -  
expClassName  
it.unitn.nlpir.experiment.fqa.<structure_generation_class_name>  
#see the next slide for the structure generator names available
```

You can use ironcreek.net/phpsyntaxtree/ to visualize the structures

how big did girl scout cookie boxes used to be

[ROOT [S [ADVP [WRB [how]]] [REL-FOCUS-QUANTITY-NP [JJ [big]] [VBD [do]] [NN [girl]] [NN [scout]] [NN [cookie]]
[REL-NNS [box]]] [VP [VBN [use]] [TO [to]] [VB [be]]]]]



Structure generators to try

- `it.unitn.nlpir.experiment.fqa.CHExperiment`
- `it.unitn.nlpir.experiment.fqa.ConstExperiment`
- `it.unitn.nlpir.experiment.fqa.DT1Experiment`
- `it.unitn.nlpir.experiment.fqa.DT2Experiment`
- `it.unitn.nlpir.experiment.fqa.LCTqDT2aExperiment`

