# Automatic question answering

## Description:

Develop the system to make it possible find answer on questions in text documents. There can be different question types. For example, if the user asks for the picture of Kyiv. Algorithm should return the link into image from text.

## Reasoning:

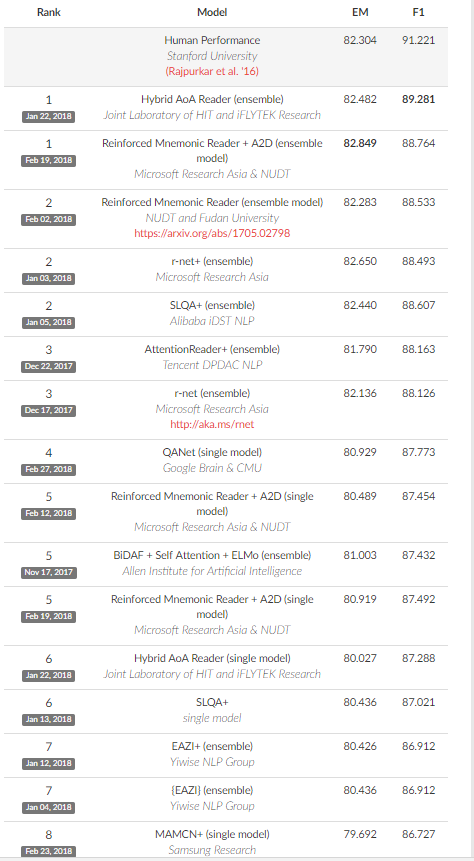
It's cool if you can easily find an answer in the big amount of documents without reading it. It's the trend in customer support, emergency fixing, information browsing. This kind of technology can significantly improve quality of medicine, transportation, logistic. In another hand, a lot of algorithms, nlp practices, ML and DL would be involved

## Steps of implementation:

1. Analize datasets: SQuAD, CNN\Dayli Mail, MSMarco, TriviaQA, WiliReading
2. Analize SoA approaches from SQuAD leaderboard:
   1. AoA https://arxiv.org/abs/1607.04423
   2. r-NET https://www.microsoft.com/en-us/research/wp-content/uploads/2017/05/r-net.pdf
   3. ReasoNet https://arxiv.org/abs/1609.05284
   4. FusionNet https://arxiv.org/abs/1711.07341
   5. HyperQA https://github.com/tjurczyk/HyperQA/blob/master/LICENSE
   6. BiDAF https://allenai.github.io/bi-att-flow/
   7. RDF Based file:///C:/Users/nazar.grycshuk/Downloads/mod490-zou.pdf
   8. Reinforcement Mnemonic Reading <https://arxiv.org/pdf/1705.02798.pdf>
3. Prepare text documents, like rules of transportation, with the topic different from step 1 dataset, this text will be target dataset for model evaluation
4. Create baseline that trains on public dataset and answers on hold
5. Implement/use one of SoA approaches (IR + Machine reading Comprehension model)
6. Build RDF based knowledge db
7. Make an ensemble from (IR + Machine reading Comprehension model )+RDF based model
8. Made tests on SQuAD, and hold dataset

# Notes

## QuAD dataset



## SoA Models

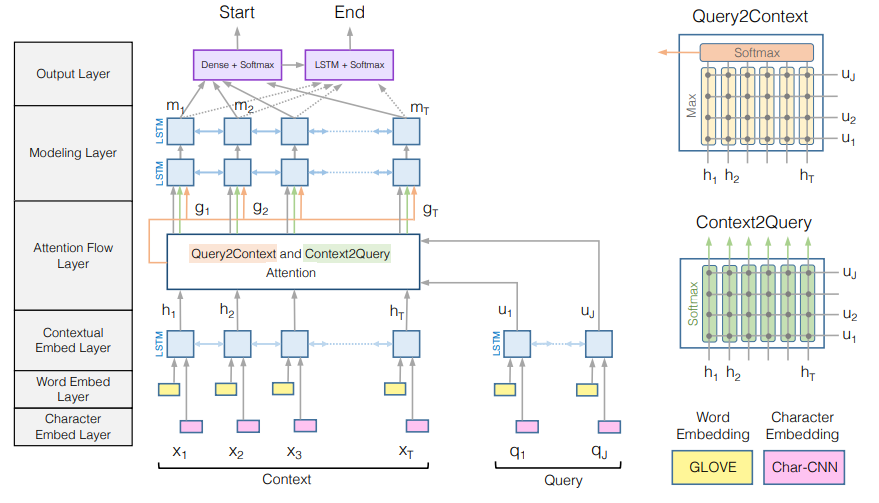
**BiDAF - Bi-Directional Attention Flow (BIDAF) network** Code on github

Attention data:

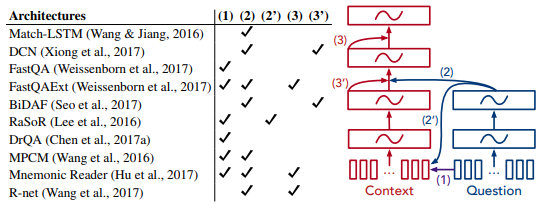
* character-level
* word-level
* contextual embeddings (Bi-LSTM on top character-level, word-level)

Attention:

* Context-to-query Attention.
* Query-to-context Attention.

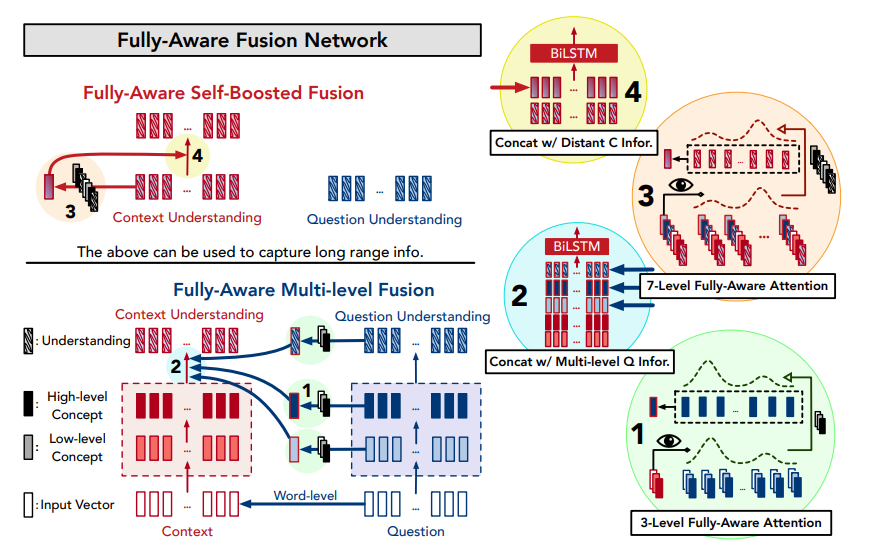


**FusionNet -** word-level embedding up to the highest semantic-level representation



**RestoNET** - Reasoning Networks

mimic the inference process of human readers Reinforcement learning



**Attention-over-Attention Neural Networks**

