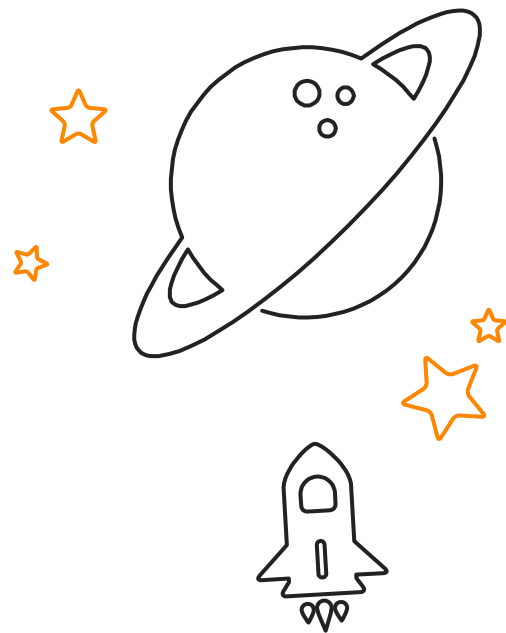


QA systems Baseline SQuAD

Профиль "Искусственный интеллект"
Олимпиады Клубного движения ИТИ
и Академии искусственного интеллекта для школьников



Stanford QA Dataset (SQuAD)

Задача SQuAD - найти ответ на вопрос в заданном контексте (например, абзаце из Википедии), где ответ на вопрос является сегментом контекста:

Контекст: В метеорологии осадки - это любой продукт конденсации атмосферного водяного пара, который падает под действием силы тяжести. Основные формы осадков включают моросящий дождь, дождь, мокрый снег, снег, крупу и град ... Осадки образуются, когда более мелкие капли сливаются в результате столкновения с другими каплями дождя или кристаллами льда в облаке. Кратковременные интенсивные периоды дождя в отдельных местах называются «ливнями».

Вопрос: Где капли воды сталкиваются с кристаллами льда, образуя осадки?

Ответ: в облаке

SQuAD2.0

The Stanford Question Answering Dataset

What is SQuAD?

Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.

SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

[Explore SQuAD2.0 and model predictions](#)

[SQuAD2.0 paper \(Rajpurkar & Jia et al. '18\)](#)

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	F1
	Human Performance Stanford University (Rajpurkar & Jia et al. '18)	86.3
1	SA-Net on Albert (ensemble) QIANXIN	90.0
2	SA-Net-V2 (ensemble) QIANXIN	90.0
2	Retro-Reader (ensemble) Shanghai Jiao Tong University http://arxiv.org/abs/2001.09694	90.0
3	ATRLP+PV (ensemble) Hithink RoyalFlush	90.0

Формат данных

Метрика: F1 / Accuracy

Формат заданий: json

paragraphs": [{ "context": "Вода, природный газ и электричество поставляются муниципалитетам кантона Женева государственной компанией Сервис-Андустриэль-ду-Женев (фр. Services Industriels de Genève). Большая часть питьевой воды (80 %) выкачивается из Женевского озера, а остальные 20 % из подземных вод, которые изначально поступают из реки Арв. 30 % электричества производится в самом кантоне тремя плотинами гидроэлектростанции на реке Рона: Сюже, Вербуа, Шанси-Пуни (фр. Seujet, Verbois, Chancy-Pougny). Кроме того, 13 % электричества производится в кантоне за счёт мусоросжигателя Ле-Шенувье (фр. Les Cheneviers). Остальные 57 % электричества поставляются из других кантонов Швейцарии и европейских стран. Государственная компания Сервис-Андустриэль-ду-Женев покупает электричество, которое произведено только за счёт методов возобновления энергии и никогда не использует электричество, произведённое на АЭС или ТЭС. Природный газ в город поставляет швейцарская компания Gaznat из Западной Европы, которая также осуществляет две трети всех поставок газа в муниципалитеты кантона Женева. Сервис-Андустриэль-ду-Женев также предоставляет телекоммуникационные услуги транспортным компаниям, поставщикам услуг, крупным предприятиям. В период с 2003 по 2005 годы работал пилотный проект Voisin, voisine (Сосед, соседка) использующий волоконно-оптический кабель с Triple play, апробированный на рынке конечных пользователей в районе Шарми (фр. Charmilles).",

"qas": [{ "id": "6c590a8747403debe65d2fd031b27f2606700b88",

"question": "Где расположены плотины гидроэлектростанции, производящей 30 % электричества в Женеве?"

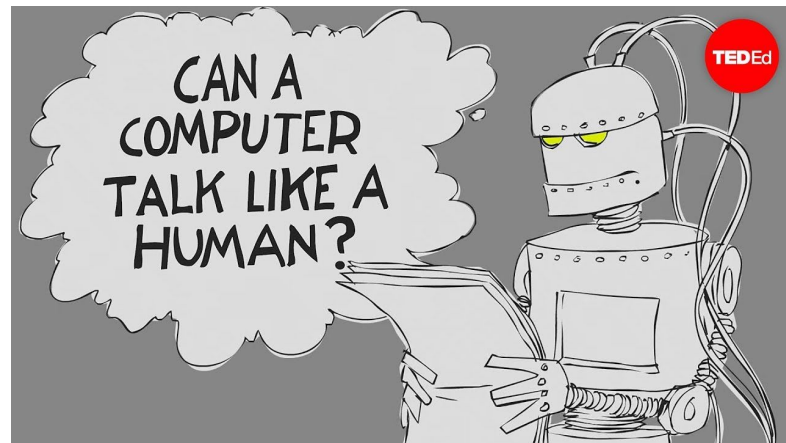
"answers": [{ "text": "на реке рона", "answer_start": 402 }],
}] }

QA-системы – какие они?

Автоматизация процессов & инженерия:

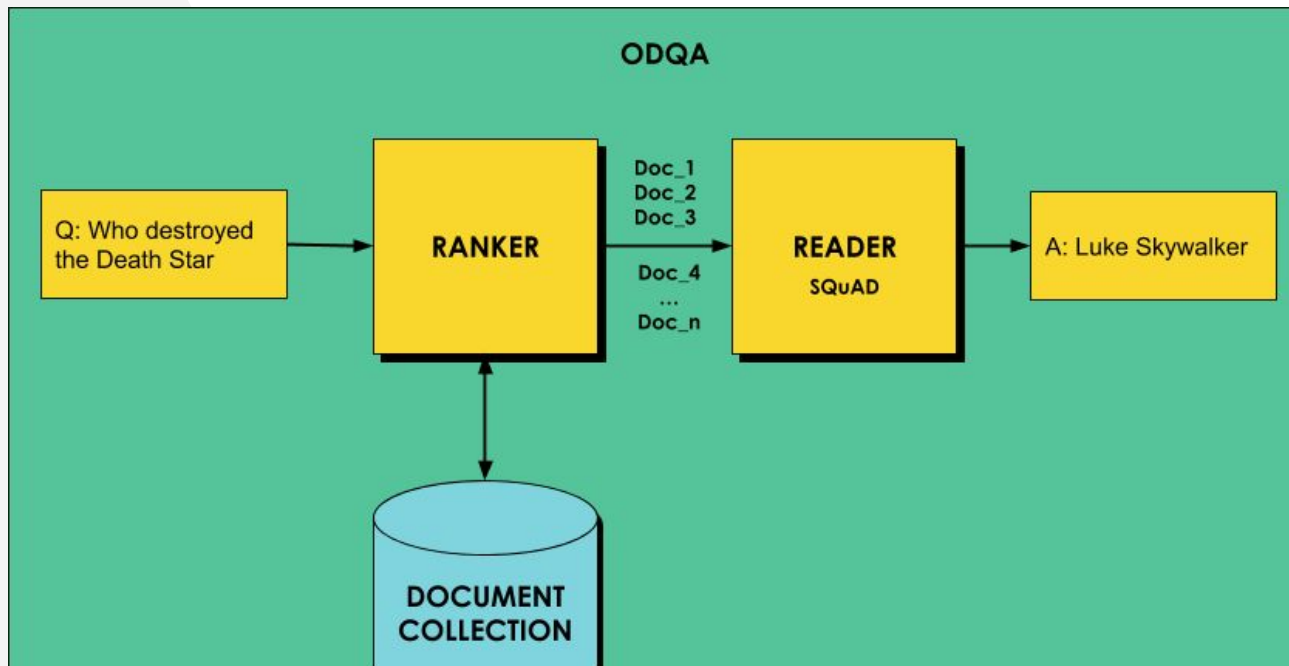
- Автоматизация колл-центров
- Обработка обратной связи от клиентов бизнеса
- Обработка обращений граждан
- Помощь в образовательных процессах

Чистая наука:



Open-Domain Question Answering (ODQA)

Ответ на вопросы (ODQA) - это задача нахождения точного ответа на любой вопрос без контекста. Модели на вход дается только вопрос, ответ нужно найти / отранжировать /



BERT SQuAD Russian

Документация <http://docs.deeppavlov.ai/en/master/features/models/squad.html>

SDSJ Task B

Pretrained models are available and can be downloaded:

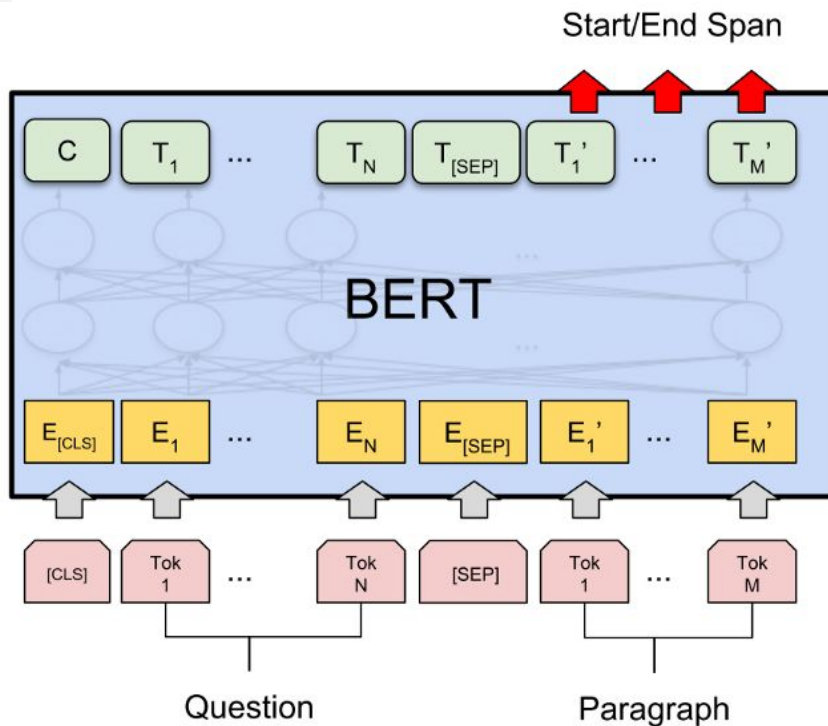
```
python -m deeppavlov download deeppavlov/configs/squad/squad_ru.json
python -m deeppavlov download deeppavlov/configs/squad/squad_ru_rubert_infer.json
python -m deeppavlov download deeppavlov/configs/squad/squad_ru_bert_infer.json
```

Link to SDSJ Task B dataset: http://files.deeppavlov.ai/datasets/sber_squad-v1.1.tar.gz

Model config	EM (dev)	F-1 (dev)
DeepPavlov RuBERT	66.30+-0.24	84.60+-0.11
DeepPavlov multilingual BERT	64.35+-0.39	83.39+-0.08
DeepPavlov R-Net	60.62	80.04

BERT SQuAD Russian

Документация <http://docs.deeppavlov.ai/en/master/features/models/squad.html>



Другие подходы

EfficientQA 2020 (NeurIPS) by Google

efficientqa.github.io/ ai.google.com/research/NaturalQuestions/efficientqa

- ▶ Natural Questions: qa-dataset with google queries, top-N results (pages) and 1 correct answer chosen by annotators
- ▶ Ограничения систем: легче 6гб, легче 500мб, самые маленькие системы с точностью не менее 25%

Другие подходы

Leaderboard results

Systems Under 6Gb

2	Model 1	FAIR-Paris&London	Facebook	11/11/2020	52.78	5.66 GB
3	Soseki	ousia_tohoku	Studio Ousia & Tohoku University	11/14/2020	50.17	5.63 GB
4	soseki	ousia_tohoku	Studio Ousia & Tohoku University	11/3/2020	50.11	4.86 GB
5	Soseki	ousia_tohoku	Studio Ousia & Tohoku University	11/7/2020	49.39	5.37 GB
6	R2D2_v2	KNOT@FIT	Brno University of Technology	11/14/2020	47.28	5.95 GB
7	R2D2	KNOT@FIT	Brno University of Technology	11/7/2020	46.06	5.38 GB

Другие подходы

Systems Under 500Mb

Rank	Model	Participant	Affiliation	Attempt Date	Accuracy	Size
1	ucInlp-fair-efficientqa-lg	ucInlp-efficientqa	University College London and Facebook AI Research	11/14/2020	33.44	336.23 MB
2	RDR-500MB	naver-clova-parsing	Naver Clova	11/14/2020	32.06	484.69 MB
3	ucInlp-fair-efficientqa-lg	ucInlp-efficientqa	University College London and Facebook AI Research	11/11/2020	29.44	226.09 MB
4	RDR-500MB	naver-clova-parsing	Naver Clova	11/9/2020	28.22	496.99 MB
5	ucInlp-fair-efficientta	ucInlp-efficientta	University College London and Facebook AI Research	11/7/2020	26.78	29.43 MB

Другие подходы

Smallest Systems Achieving 25% Accuracy

Rank	Model	Participant	Affiliation	Attempt Date	Accuracy	Size
1	ucInlp-fair-efficientqa	ucInlp-efficientqa	University College London and Facebook AI Research	11/7/2020	26.78	29.43 MB
2	ucInlp-fair-efficientqa	ucInlp-efficientqa	University College London and Facebook AI Research	10/27/2020	25.78	157.26 MB
3	ucInlp-fair-efficientqa-lg	ucInlp-efficientqa	University College London and Facebook AI Research	11/11/2020	29.44	226.09 MB
4	ucInlp-fair-efficientqa-lg	ucInlp-efficientqa	University College London and Facebook AI Research	11/14/2020	33.44	336.23 MB
5	RDR-500MB	naver-clova-narsina	Naver Clova	11/14/2020	32.06	484.69 MB

Некоторые полезные ссылки:

Репозиторий с сегодняшним ноутбуком:

https://github.com/AI-Front/NTI/blob/main/Train_SQuaD_tutorial.ipynb

Отчет организаторов EfficientQA с описанием систем:

<https://arxiv.org/abs/2101.00133>