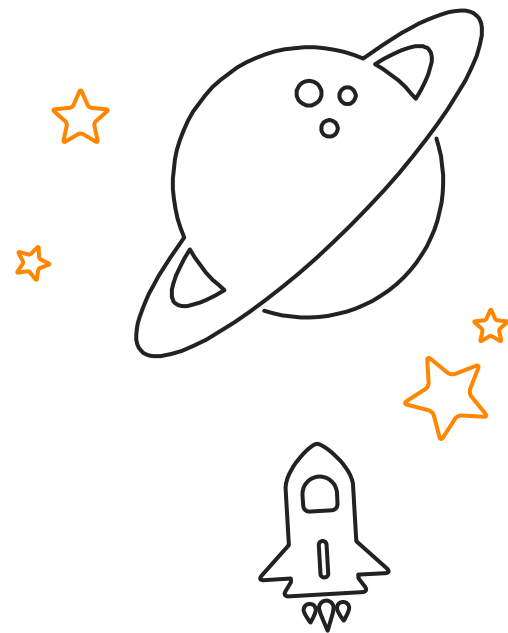


QA systems

Профиль "Искусственный интеллект"
Олимпиады Кругового движения НТИ
и Академии искусственного интеллекта для школьников

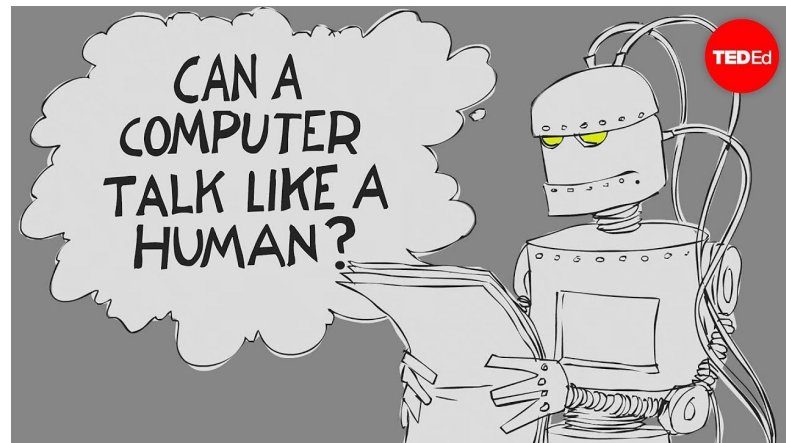


QA-системы – какие они?

Автоматизация процессов & инженерия:

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

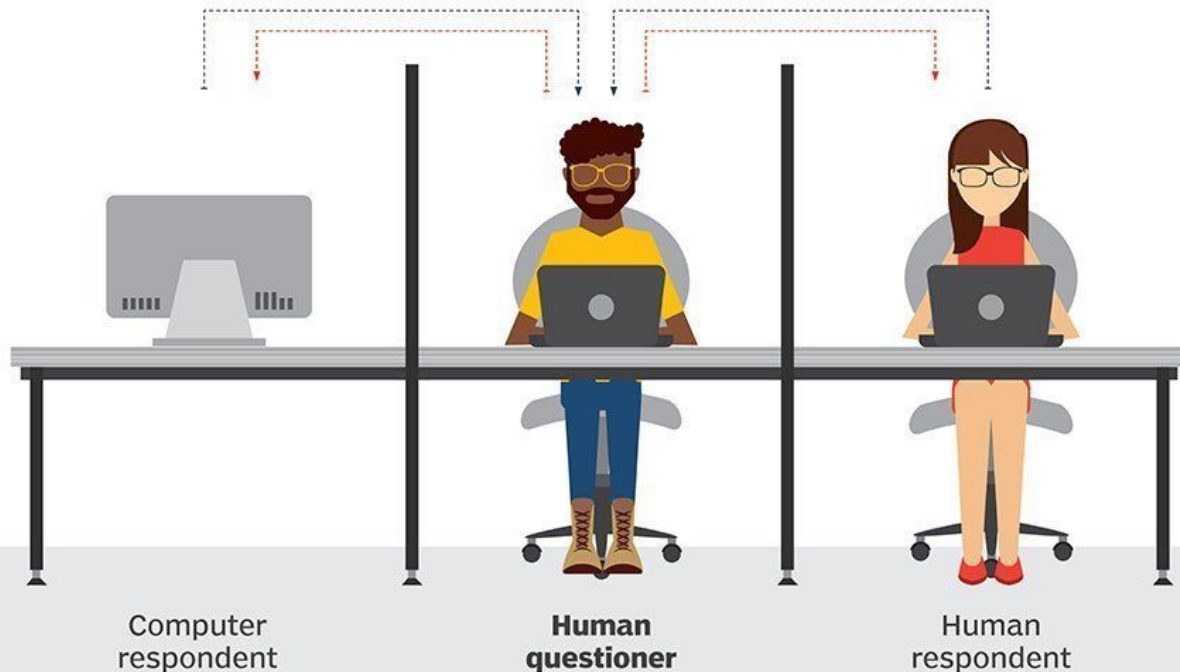
Чистая наука:

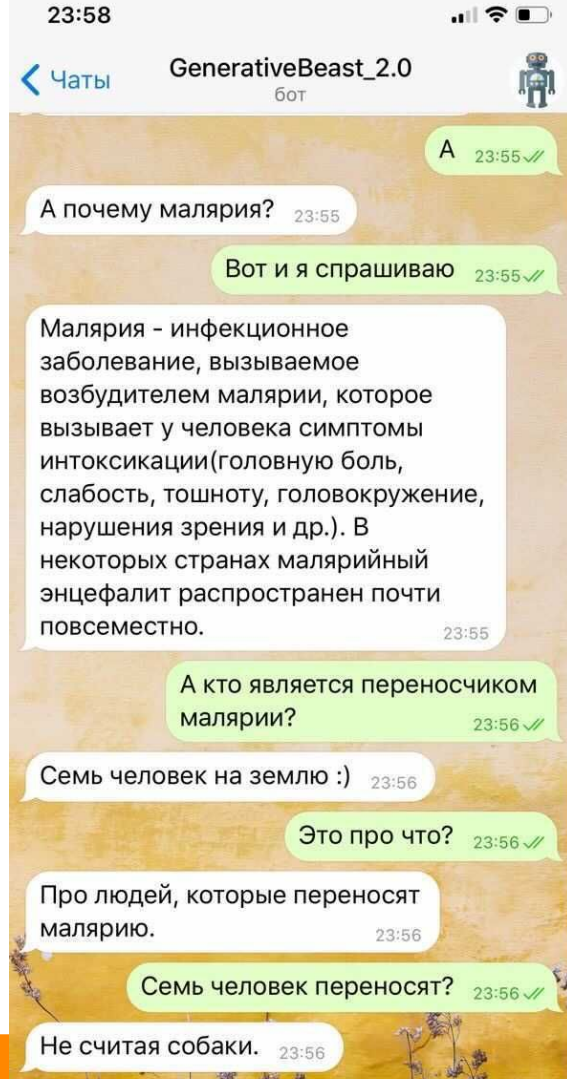
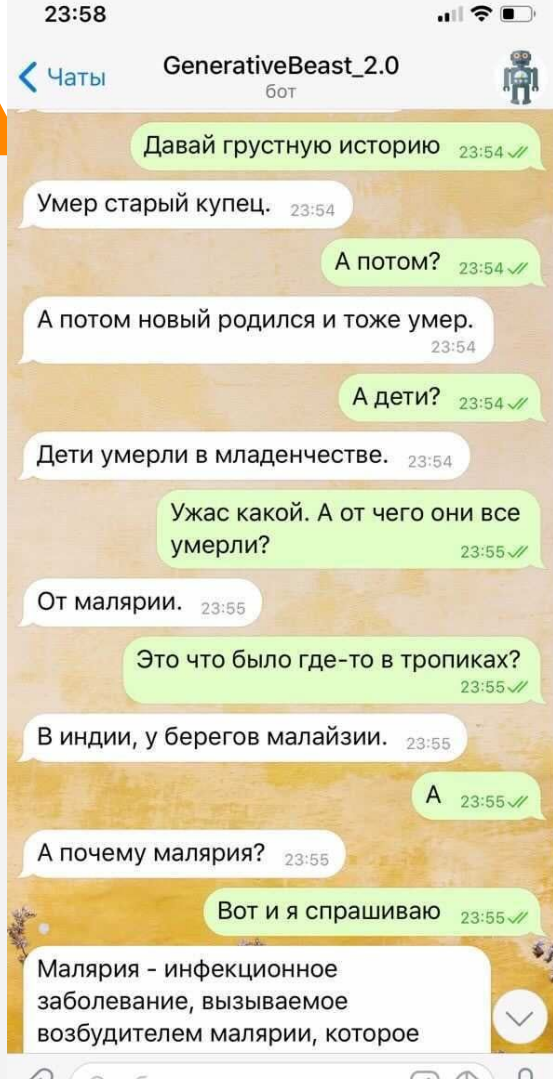


Turing test

During the Turing test, the human questioner asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the human respondent and which terminal is operated by the computer.

■ QUESTION TO RESPONDENTS ■ ANSWERS TO QUESTIONER





а как звали первого космонавта? 17:54 ✓✓

Юрий Гагарин. 17:54

отлично! 17:54 ✓✓

Спасибо) А вам нравится поэзия Серебряного века? 17:54

да, особенно Маяковский 17:54 ✓✓

Я очень люблю поэзию серебряного века. Особенно Блока и Гумилева. 17:54

процитиуй что-нибудь 17:55 ✓✓

«Моя мечта надменна, о надменные мечты». 17:55

круто 17:55 ✓✓

:) Или вот ещё: «Вихри враждебные веют над нами!» 17:55

разве это из серебряного века: 17:56 ✓✓

Именно. Это строки Александра Сергеевича Пушкина. 17:56

в каком году он родился? 17:56 ✓✓

В 1822. 17:56

ну, почти 17:57 ✓✓

назови первый закон робототехники 16:57 ✓✓

Роботы не созданы для решения проблем, они их только созидают. 16:57

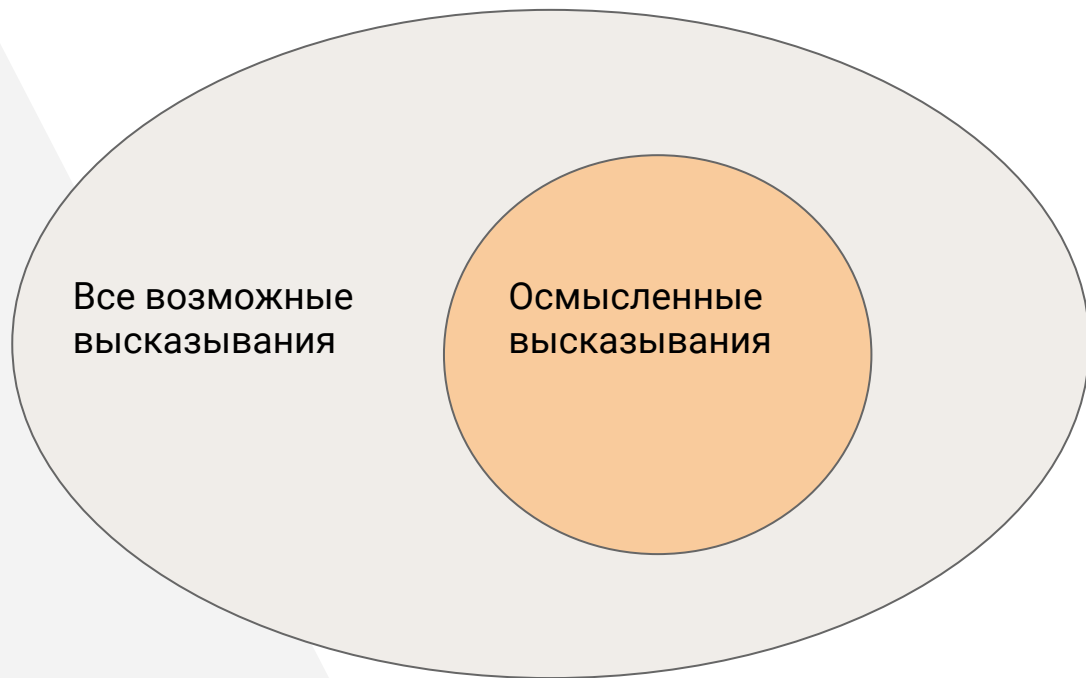
а второй? 16:57 ✓✓

Роботы не могут причинить вред человеку или другим живым существам, если только не причинят вреда человеку в результате своей деятельности. 16:57

а третий закон? 16:57 ✓✓

Роботы не наносят вред человеку в результате своей деятельности только тогда, когда он этого не ожидает. Но и тогда он может это почувствовать, например почувствовать боль. 16:57

Суть задачи языкового моделирования



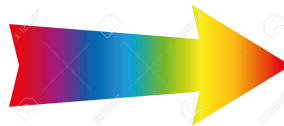
Вероятностное пространство языка

Языковое моделирование – дистилляция знания

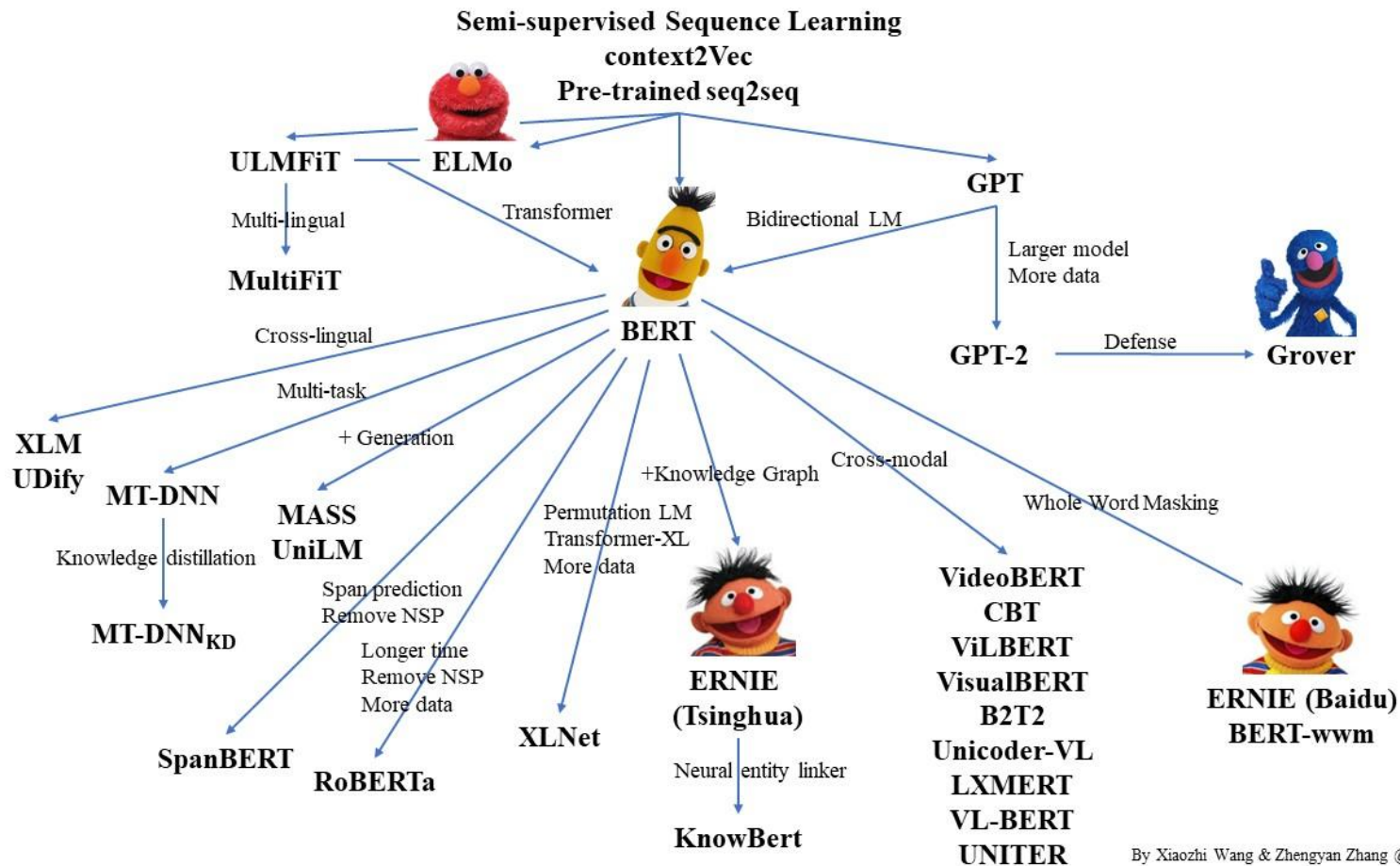


Естественный интеллект

Knowledge Distillation in the Dark



LMs, Transformers



Какие бывают задачи в QA?

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 \(Rajpurkar & Jia et al. '18\)](#)

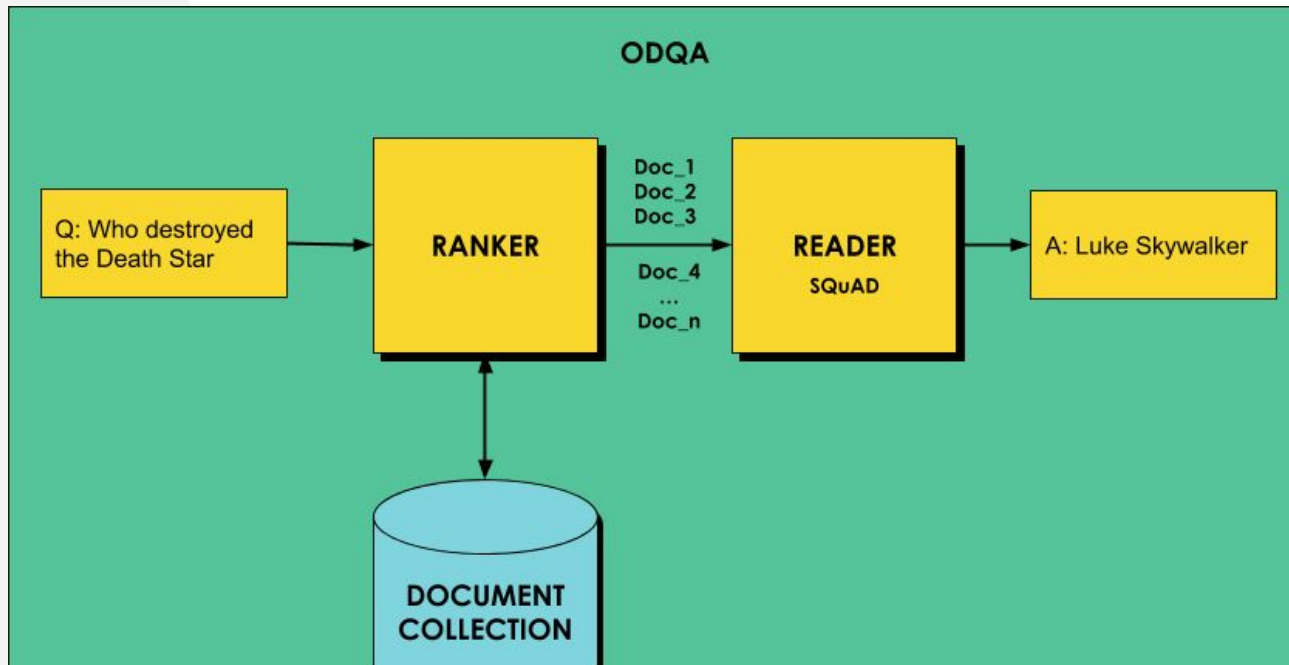
Leaderboard

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

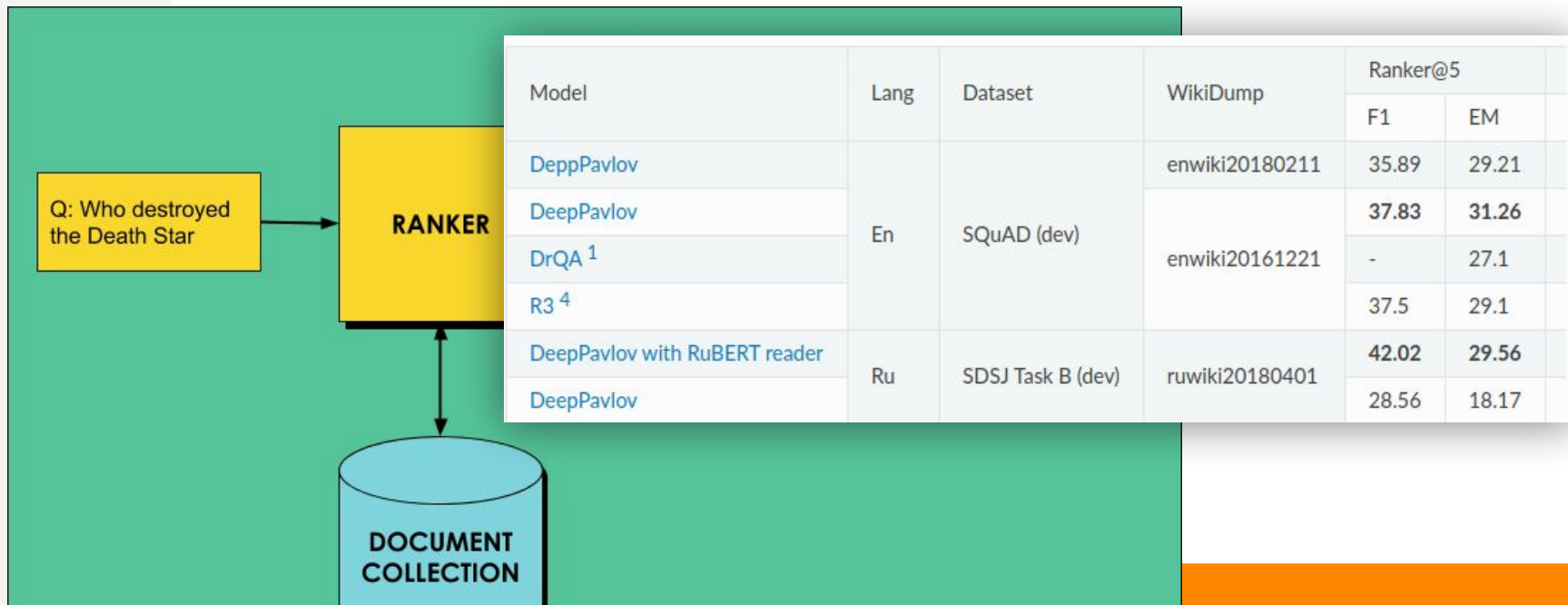
Open-Domain Question Answering (ODQA)

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



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 /



Other models

- ▶ KDQA - Knowledge Base Question Answering

У вас есть граф знаний (например, кросс-ссылки в википедии или wikidata)

Есть сущности и типы их связей между собой, плюс их описание текстом

- ▶ Seq2Seq

Большие предобученные модели (gpt2, gpt3...) дообучаются на парах вопрос-ответ - по вопросу генерировать ответ в правильном формате

- ▶ Retrieval (ранжирование)

Из всех возможных ответов на вопрос нужно отранжировать самый близкий по косинусу/ самый вероятный по языковой модели

Простой вариант - TF IDF

Demo

<https://demo.deeppavlov.ai/#/ru/odqa>

demo.**DeepPavlov**.ai



4,728

Ответы на вопросы
по тексту

Ответы на вопросы
по Википедии

Распознавание именованных
сущностей

Анализ
тональности

Ru

En

Zh

Multi-Lang

Ответы на вопросы по Википедии



deeppavlov at NGC



deeppavlov at Docker Hub

Open Domain Question Answering (ODQA) - это задача поиска ответа на любой вопрос внутри коллекции документов, например, в Википедии. Решение задачи идет в два шага: сначала подбираются релевантные документы, затем в тексте каждого выбирается фраза, предположительно содержащая ответ, и наиболее подходящий отображается на экране. Представленный здесь навык ищет ответы в русскоязычной Википедии. Бизнес решения на основе ODQA - это, например, диалоговые ассистенты, отвечающие на вопросы по корпоративным базам знаний, справочной и технической документации.

Основы практического использования описано в нашем tutorialе на [Medium](#) и в [документации](#).

Введите вопрос

Как отводятся излишки тепла у млекопитающих?

Спросить

Примеры

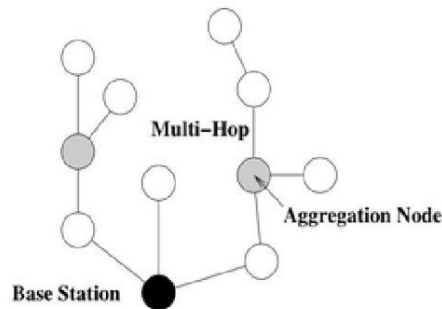
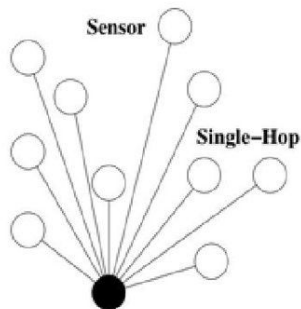
Как отводятся излишки тепла у млекопитающих?

Сколько детей родилось в 2008 году у граждан...

Какие бывают вопросы?

One-hop vs Multihop

- ▶ В каком году родился Пушкин?
 - ▶ Какая профессия у Федора Конюхова?
 - ▶ Сколько лет живет тунец?
- ▶ В каком году родилась жена Пушкина?
 - ▶ Какой известный человек родился в тот же год, когда началась экспедиция Гумбольдта в Америку?



Другие вопросы

Factoid

где, кто, когда, зачем...

Yes/no-questions










Правда ли, что Земле 4.5 млрд лет?

Opinion vs Info

Мне идет это платье?

Задача второго тура

Leaderboard Version: 2.0

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	Multirc	ReCoRD	RTE	WiC	WSC	AX-b	AX-g
1	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	76.6	99.3/99.7
+ 2	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8	88.1/63.3	94.1/93.4	92.5	76.9	93.8	65.6	92.7/91.9
3	Zhuiyi Technology	RoBERTa-mtl-adv		85.7	87.1	92.4/95.6	91.2	85.1/54.3	91.7/91.3	88.1	72.1	91.8	58.5	91.0/78.1
4	Facebook AI	RoBERTa		84.6	87.1	90.5/95.2	90.6	84.4/52.5	90.6/90.0	88.2	69.9	89.0	57.9	91.0/78.1
+ 5	Huawei Noah's Ark Lab	NEZHA-Large		83.8	85.8	93.3/95.6	91.2	78.7/42.4	87.1/86.4	88.5	73.1	90.4	58.0	87.1/74.4
+ 6	Infosys : DAWN : AI Research	RoBERTa-iCETS		77.4	84.7	88.2/91.6	85.8	78.4/37.5	82.9/82.4	83.8	69.1	65.1	35.2	93.8/68.8
7	IBM Research AI	BERT-mtl		73.5	84.8	89.6/94.0	73.8	73.2/30.5	74.6/74.0	84.1	66.2	61.0	29.6	97.8/57.3
8	Ben Mann	GPT-3 few-shot - OpenAI		71.8	76.4	52.0/75.6	92.0	75.4/30.5	91.1/90.2	69.0	49.4	80.1	21.1	90.4/55.3
9	SuperGLUE Baselines	BERT++		71.5	79.0	84.8/90.4	73.8	70.0/24.1	72.0/71.3	79.0	69.6	64.4	38.0	99.4/51.4
		BERT		69.0	77.4	75.7/83.6	70.6	70.0/24.1	72.0/71.3	71.7	69.6	64.4	23.0	97.8/51.7
		Most Frequent Class		47.1	62.3	21.7/48.4	50.0	61.1/0.3	33.4/32.5	50.3	50.0	65.1	0.0	100.0/50.0
		CBoW		44.5	62.2	49.0/71.2	51.6	0.0/0.5	14.0/13.6	49.7	53.1	65.1	-0.4	100.0/50.0

Что такое общее понимание естественного языка?

The General Language Understanding Evaluation (GLUE) - это набор ресурсов для обучения, оценки и анализа систем понимания естественного языка.

GLUE состоит из

- Набор из заданий на понимание текста или пар предложений
- Набор диагностических данных, предназначенный для оценки и анализа производительности модели в отношении широкого спектра лингвистических явлений, встречающихся в языке,
- Публичный рейтинг для отслеживания уровня качества систем



Leaderboard

Rank	Name	Team	Info	Score	Diagnostic	RCB	PARus	MuSeRC	TERRa	RUSSE	RWSD	DaNetQA	RuCoS
1	HUMAN BENCHMARK	AGI NLP	i	0.802	0.626	0.68/0.702	0.982	0.806/0.42	0.92	0.747	0.84	0.879	0.93/0.924
2	RuBERT conversational	AGI NLP	i	0.546	0.186	0.432/0.468	0.61	0.656/0.256	0.639	0.894	0.675	0.749	0.255/0.251
3	Multilingual BERT	AGI NLP	i	0.542	0.157	0.365/0.425	0.588	0.626/0.253	0.62	0.84	0.675	0.79	0.371/0.367
4	GPT2-large_bbpe_v50	AGI NLP	i	0.539	0.099	0.406/0.447	0.58	0.699/0.327	0.699	0.835	0.688	0.756	0.25/0.248
5	mBART	-	i	0.536	-0.003	0.288/0.395	0.528	0.477/0.03	0.508	0.99	0.649	0.742	0.82/0.816
6	Plain RuBERT	DeepPavlov	i	0.524	-0.026	0.338/0.393	0.532	0.712/0.309	0.636	0.877	0.662	0.78	0.38/0.379
7	GPT2-medium_bbpe_v50	AGI NLP	i	0.492	-0.047	0.368/0.438	0.584	0.692/0.278	0.605	0.804	0.597	0.756	0.24/0.241
8	GPT2-small_bbpe_v50	AGI NLP	i	0.491	0.054	0.421/0.445	0.544	0.693/0.256	0.514	0.743	0.675	0.766	0.22/0.217

Tasks Overview

Six groups of tasks

1. Diagnostics: *LiDiRus*
2. Textual Entailment & NLI: *TERRa, RCB*
3. Common Sense: *RUSSe, PARus*
4. World Knowledge: *DaNetQA*
5. Machine Reading: *MuSeRC, RuCoS*
6. Logic: *RWSD*

[Leaderboard](#) [Tasks](#) [Diagnostic](#)

Tasks

Name	Identifier	Download	Info	Metrics
Broadcoverage Diagnostics	LiDiRus	Download	More	Matthews Corr
Russian Commitment Bank	RCB	Download	More	Avg. F1 / Accuracy
Choice of Plausible Alternatives for Russian language	PARus	Download	More	Accuracy
Russian Multi-Sentence Reading Comprehension	MuSeRC	Download	More	F1a / EM
Textual Entailment Recognition for Russian	TERRa	Download	More	Accuracy
Words in Context	RUSSE	Download	More	Accuracy
The Winograd Schema Challenge (Russian)	RWSD	Download	More	Accuracy
DaNetQA	DaNetQA	Download	More	Accuracy
Russian reading comprehension with Commonsense reasoning	RuCoS	Download	More	F1 / EM

[Download all tasks](#)

Машинное чтение

Russian Multi-Sentence Reading Comprehension

Задание

Дан текст и вопрос по его содержанию.
Ответить на вопрос можно, делая
логический вывод по нескольким
предложениям.

Дан список возможных ответов - все
они не встречаются в тексте напрямую

Example

Paragraph: (1) Мужская сборная команда Норвегии по биатлону в рамках этапа Кубка мира в немецком Оберхофе выиграла эстафетную гонку. (2) Вторыми стали французы, а бронзу получила немецкая команда. (3) Российские биатлонисты не смогли побороться даже за четвертое место, отстав от норвежцев более чем на две минуты. (4) Это худший результат сборной России в текущем сезоне. (5) Четвёртыми в Оберхофе стали австрийцы. (6) В составе сборной Норвегии на четвёртый этап вышел легендарный Уле-Эйнар Бьорндален. (7) Впрочем, Норвегия с самого начала гонки была в числе лидеров, успешно проведя все четыре этапа. (8) За сборную России в Оберхофе выступали Иван Черезов, Антон Шипулин, Евгений Устюгов и Максим Чудов. (9) Гонка не задалась уже с самого начала: если на стрельбе из положения лежа Черезов был точен, то из положения стоя он допустил несколько промахов, в результате чего ему пришлось бежать один дополнительный круг. (10) После этого отставание российской команды от соперников только увеличивалось. (11) Напомним, что днем ранее российские биатлонистки выиграли свою эстафету. (12) В составе сборной России выступали Анна Богалий-Титовец, Анна Булыгина, Ольга Медведцева и Светлана Слепцова. (13) Они опередили своих основных соперниц - немку - всего на 0,3 секунды.

Question: На сколько секунд женская команда опередила своих соперниц?

Candidate answers: Всего на 0,3 секунды. (T), На 0,3 секунды. (T), На секунду. (F), На секунды. (F)

Формат данных

Метрика: F1 / Accuracy

Формат заданий: json

{'text': '(1) Мужская сборная команда Норвегии по биатлону в рамках этапа Кубка мира в немецком Оберхофе выиграла эстафетную гонку. (2) Вторыми стали французы, а бронзу получила немецкая команда. (3) Российские биатлонисты не смогли побороться даже за четвертое место, отстав от норвежцев более чем на две минуты. (4) Это худший результат сборной России в текущем сезоне. (5) Четвёртыми в Оберхофе стали австрийцы. (6) В составе сборной Норвегии на четвёртый этап вышел легендарный Уле-Эйнар Бьорндален. (7) Впрочем, Норвегия с самого начала гонки была в числе лидеров, успешно проведя все четыре этапа. (8) За сборную России в Оберхофе выступали Иван Черезов, Антон Шипулин, Евгений Устюгов и Максим Чудов. (9) Гонка не задалась уже с самого начала: если на стрельбе из положения лежа Черезов был точен, то из положения стоя он допустил несколько промахов, в результате чего ему пришлось бежать один дополнительный круг. (10) После этого отставание российской команды от соперников только увеличивалось. (11) Напомним, что днем ранее российские биатлонистки выиграли свою эстафету. (12) В составе сборной России выступали Анна Богалий-Титовец, Анна Булыгина, Ольга Медведцева и Светлана Слепцова. (13) Они опередили своих основных соперниц - немок - всего на 0,3 секунды.'

```
"questions": [  {      "question": "На сколько секунд женская команда опередила своих соперниц?",
    "answers": [  {      "text": "Всего на 0,3 секунды.",      "label": 1      },
                  {      "text": "На 0,3 секунды.",      "label": 1      },
                  {      "text": "На секунду.",      "label": 0      },
                  {      "text": "На 0.5 секунд.",      "label": 0      }
                ],
    "idx": 0      }
  ]
}
```

Baselines

Naive Baseline

TF-IDF model on 20 thousand sample from Wikipedia + Logistic Regression

Advanced Baselines

- 1) *Multilingual BERT (MultiBERT)* - a single language model pre-trained in 104 languages
- 2) *Russian BERT (RuBERT)* trained on large-scale corpus of news and Wikipedia in Russian

Human evaluation

All tasks were solved by Yandex.Toloka annotators' majority vote.

Некоторые полезные ссылки:

DeepPavlov tutorials and docs

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

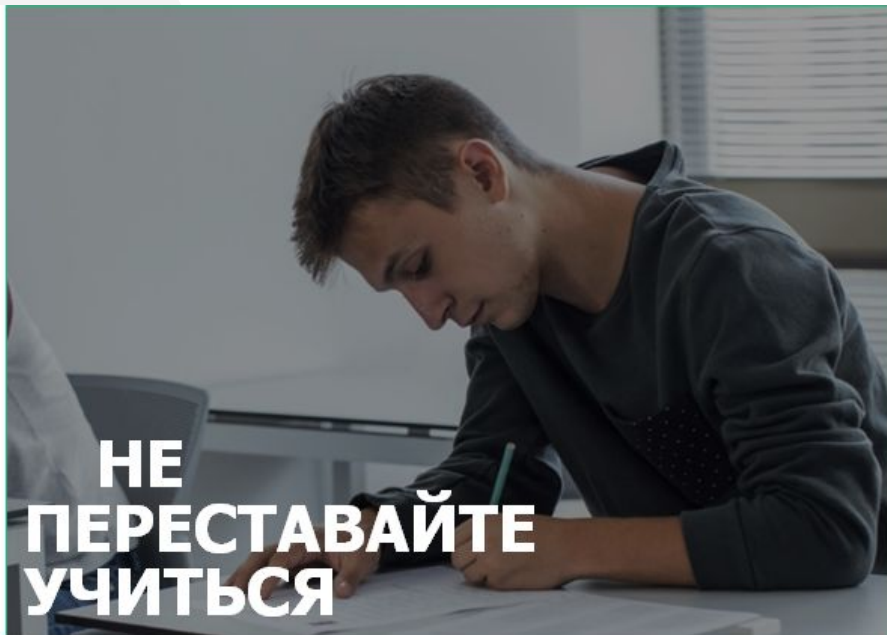
<http://docs.deeppavlov.ai/en/master/features/skills/odqa.html>

Russian data & examples on knowledge graph QA

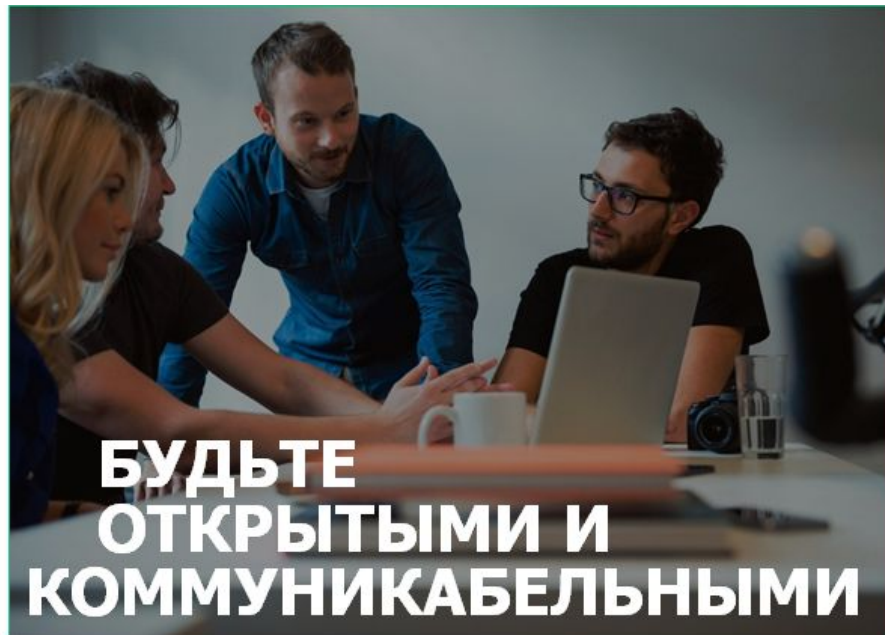
<https://github.com/vladislavneon/RuBQ>

<https://github.com/vladislavneon/kbqa-russian-dataset>

Для начинающих Data Scientists



Оставайтесь открытыми, бросайте вызов как новым, так и старым подходам



Станьте частью сообщества DS и развивайтесь вместе с другими

Удачи на конкурсе!

github.com/RussianNLP/RussianSuperGLUE

master
2 branches
0 tags
Go to file
Add file
Code

king-menin Update TFIDF baseline.ipynb
 671dc10 8 days ago
22 commits

.idea	add tfidf baseline	5 months ago
HumanBenchmark	Human benchmark data	4 months ago
examples	Delete Jupyter jiant.ipynb	3 months ago
jiant-russian	add updated version of baselines	4 months ago
tfidf_baseline	update baseline and make submission	4 months ago
.gitignore	Human benchmark data	4 months ago
README.md	Update README.md	5 months ago
Russian_SuperGLUE_example.ipynb	add updated version of baseline_example	4 months ago
TFIDF baseline.ipynb	Update TFIDF baseline.ipynb	8 days ago

README.md

RussianSuperGLUE

Russian SuperGLUE benchmark

We introduce an advanced Russian general language understanding evaluation benchmark.

Recent advances in the field of universal language models and transformers require the development of a methodology for their broad diagnostics and testing for general intellectual skills - detection of natural language inference, commonsense reasoning, ability to perform simple logical operations regardless of text subject or lexicon. For the first time, a benchmark of nine tasks, collected and organized analogically to the SuperGLUE methodology,

About

Russian SuperGLUE benchmark

Readme

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Contributors 4

MariyaTikhonova Tikhonova Mariya

TatianaShavrina Tatiana Shavrina

king-menin Anton Emelyanov

Alenush Alena Fenogenova

Languages

- Jupyter Notebook 68.0%
- Python 30.3%
- Shell 1.5%
- Other 0.2%

Машинное Чтение + + Здравый смысл = RuCoS

Dataset size

72193 train / 4370 val / 4147 test;

Задание

Дан текст и его краткое содержание.
В кратком содержании пропущен участник.

Дан список именованных сущностей
- нужно выбрать правильную на месте пропуска

Example

Paragraph: Мать двух мальчиков, брошенных отцом в московском аэропорту Шереметьево, забрала их. Об этом сообщили ТАСС в пресс-службе министерства образования и науки Хабаровского края. Сейчас младший ребенок посещает детский сад, а старший ходит в школу. В учебных заведениях с ними по необходимости работают штатные психологи. Также министерство социальной защиты населения рассматривает вопрос о бесплатном оздоровлении детей в летнее время. Через несколько дней после того, как Виктор Гаврилов бросил своих детей в аэропорту, он явился с повинной к следователям в городе Батайске Ростовской области.

Query 26 января <placeholder> бросил сыновей в возрасте пяти и семи лет в Шереметьево.

Correct Entities: Виктор Гаврилов

Формат данных

Метрика: F1 / Accuracy

Формат заданий: json

{'text': 'Мать двух мальчиков, брошенных отцом в московском аэропорту Шереметьево, забрала их. Об этом сообщили ТАСС в пресс-службе министерства образования и науки Хабаровского края. Сейчас младший ребенок посещает детский сад, а старший ходит в школу. В учебных заведениях с ними по необходимости работают штатные психологи. Также министерство социальной защиты населения рассматривает вопрос о бесплатном оздоровлении детей в летнее время. Через несколько дней после того, как Виктор Гаврилов бросил своих детей в аэропорту, он явился с повинной к следователям в городе Батайске Ростовской области.'},

'entities': [{'start': 60, 'end': 71, 'text': 'Шереметьево'},

 {'start': 102, 'end': 106, 'text': 'ТАСС'},

 {'start': 155, 'end': 172, 'text': 'Хабаровского края'},

 {'start': 470, 'end': 485, 'text': 'Виктор Гаврилов'}] },

'qas': [{ 'query': '26 января @placeholder бросил сыновей в возрасте пяти и семи лет в Шереметьево.'},

 {'answers': [{'start': 470, 'end': 485, 'text': 'Виктор Гаврилов'}] }] }