

# QuestionAnswering



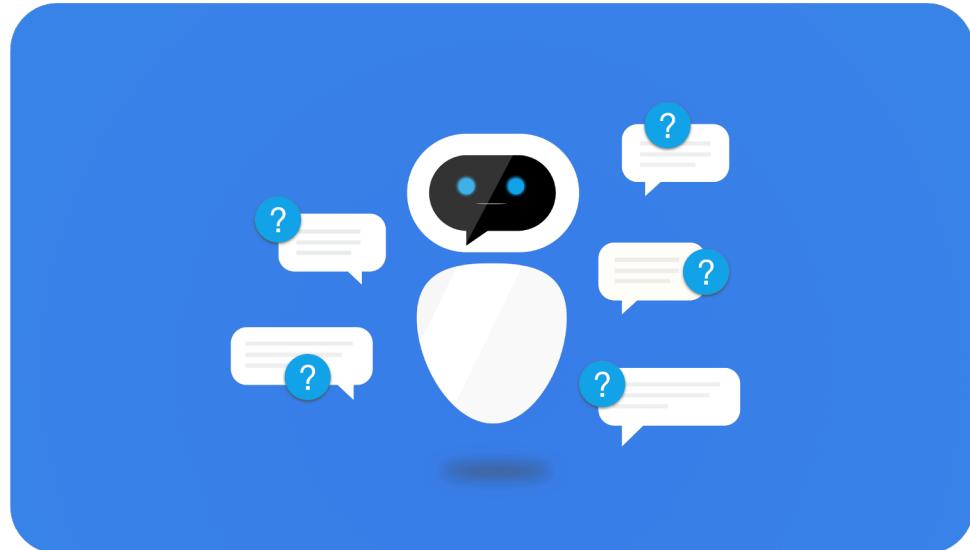
HSE  
Maria Tikhonova  
 @mashkka\_ds  
2023

Based on

<https://github.com/king-menin/mipt-nlp2022>

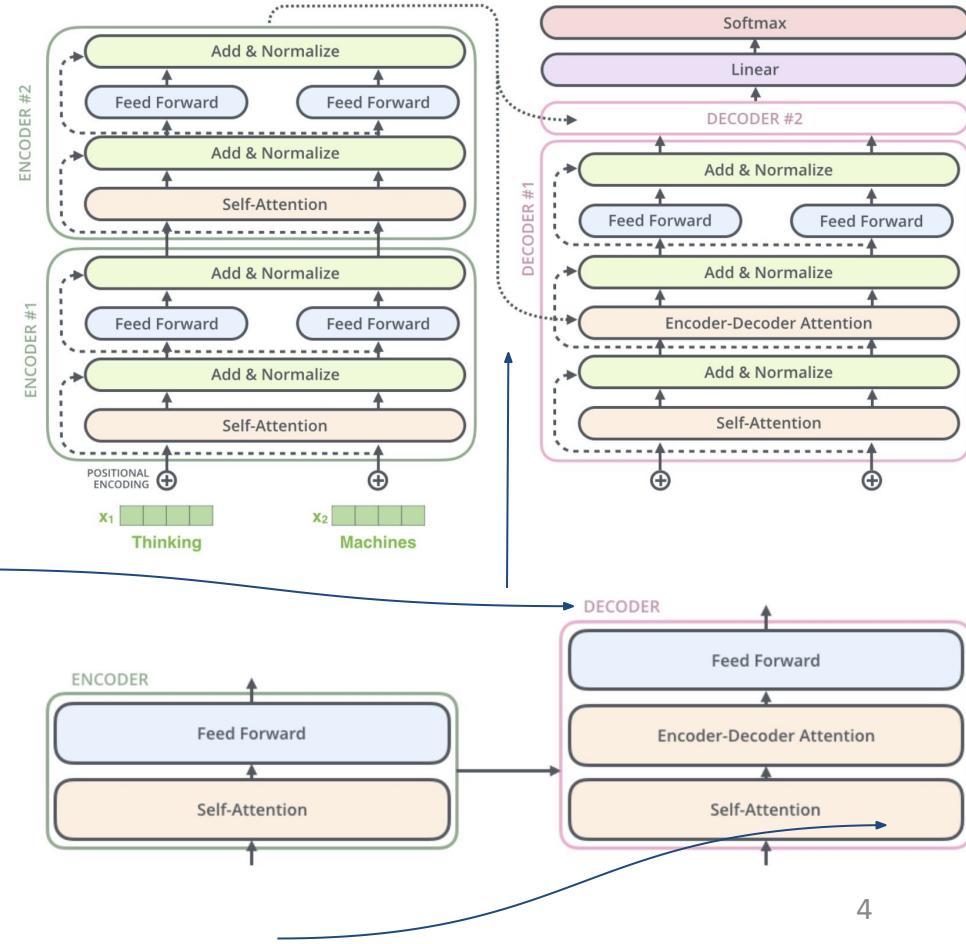
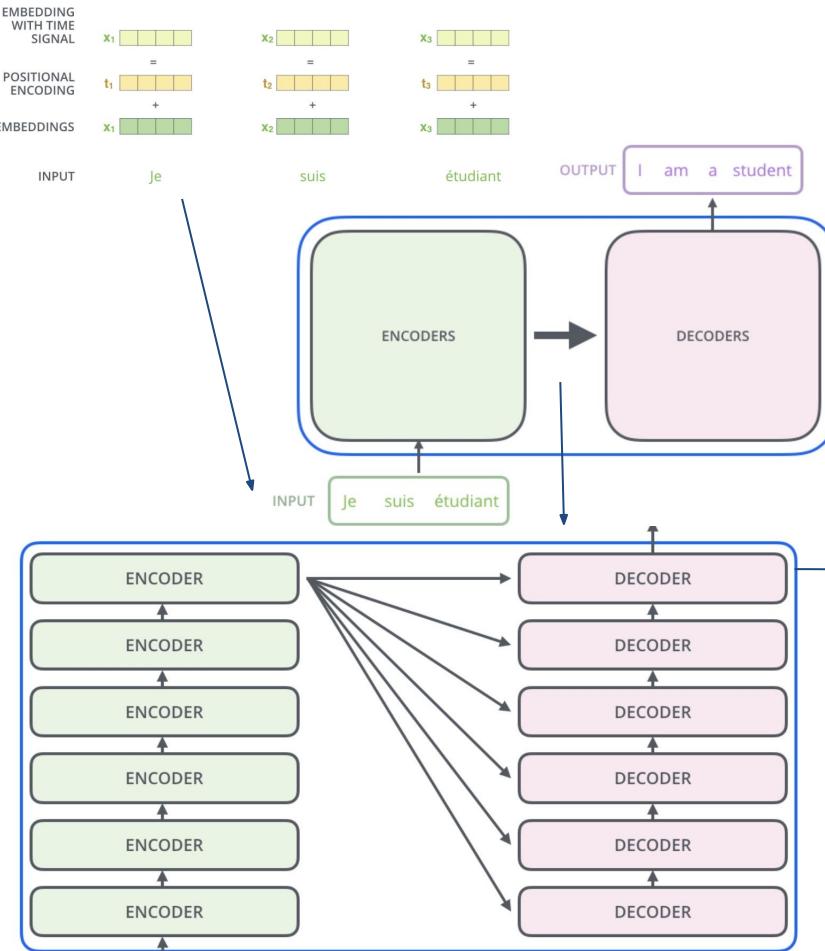
# Today

- Recap: Transformers
- What is question answering?  
Reading comprehension
- Open domain question answering
- KBQA
- Chatbots



# Recap: Transformers

# Transformers. High-level



# Self-attention

•

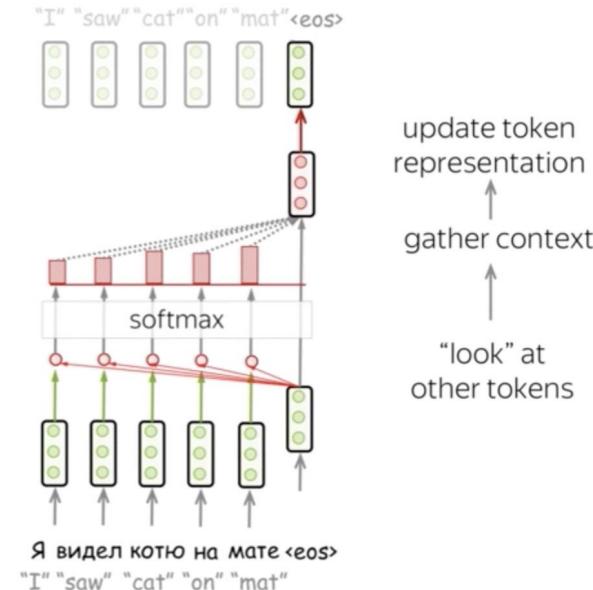
## Self-attention:

- токены взаимодействуют между собой, каждый токен «смотрит» на соседей
- собираем информацию о контексте
- обновляем векторное представление токена

Tokens try to understand themselves better in context of each other

Layer  $k + 1$

Layer  $k$



И все это параллельно!

# Query, Key and Value

## Query, Key and Value vectors

Каждый токен кодируется тремя векторами:

- **query** – запрашиваем информацию
- **key** – говорим, что у токена есть полезная информация
- **value** – отдаём информацию

Механизм attention сравнивает, «подходит ли «query и key», если да, то отдаёт информацию, содержащуюся в векторе value

$$\begin{matrix} \mathbf{x} \\ \times \\ \mathbf{W^Q} \end{matrix} = \mathbf{Q}$$

$$\begin{matrix} \mathbf{x} \\ \times \\ \mathbf{W^K} \end{matrix} = \mathbf{K}$$

$$\begin{matrix} \mathbf{x} \\ \times \\ \mathbf{W^V} \end{matrix} = \mathbf{V}$$



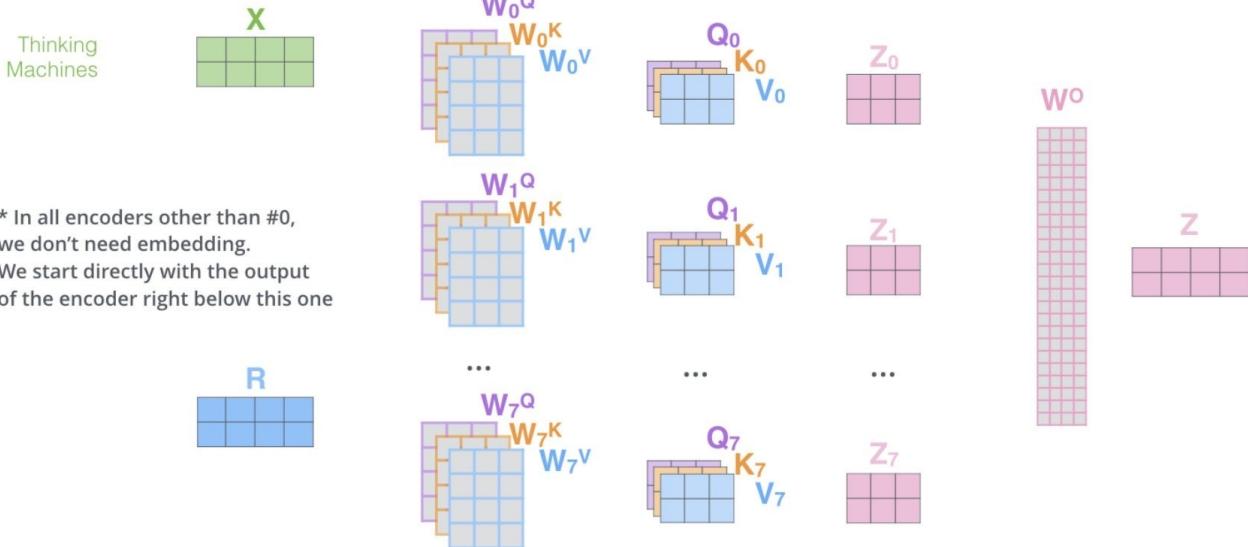
Attention weights

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}$$

# Multi-head attention

ХОТИМ ВЫУЧИВАТЬ разную  
информацию о  
предложении

- 1) This is our input sentence\*
- 2) We embed each word\*
- 3) Split into 8 heads. We multiply  $X$  or  $R$  with weight matrices
- 4) Calculate attention using the resulting  $Q/K/V$  matrices
- 5) Concatenate the resulting  $Z$  matrices, then multiply with weight matrix  $W^O$  to produce the output of the layer

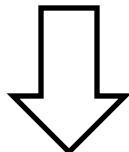


$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O$$

$$\text{where } \text{head}_i = \text{Attention}\left(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V\right)$$

# Multi-head attention

ХОТИМ ВЫУЧИВАТЬ разную  
информацию о  
предложении



Создаем несколько «голов»  
attention (multi-head  
attention), которые  
работают независимо и  
фокусируются на разных  
аспектах

- 1) This is our input sentence\*
- 2) We embed each word\*
- 3) Split into 8 heads. We multiply  $X$  or  $R$  with weight matrices
- 4) Calculate attention using the resulting  $Q/K/V$  matrices
- 5) Concatenate the resulting  $Z$  matrices, then multiply with weight matrix  $W^O$  to produce the output of the layer

Thinking  
Machines

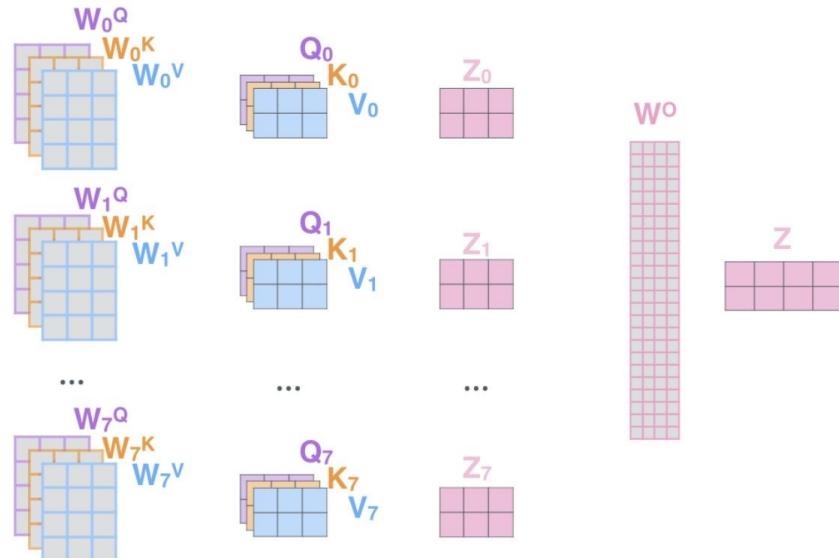


\* In all encoders other than #0, we don't need embedding. We start directly with the output of the encoder right below this one



$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O$$

$$\text{where } \text{head}_i = \text{Attention}\left(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V\right)$$



## Feed-forward blocks:

Каждый слой содержит feed-forward блок - 2 линейных слоя с ReLU между ними

$$FFN(x) = \max(0, xW_1 + b_1)W_2 + b_2.$$

## Residual connection для улучшения обучения:

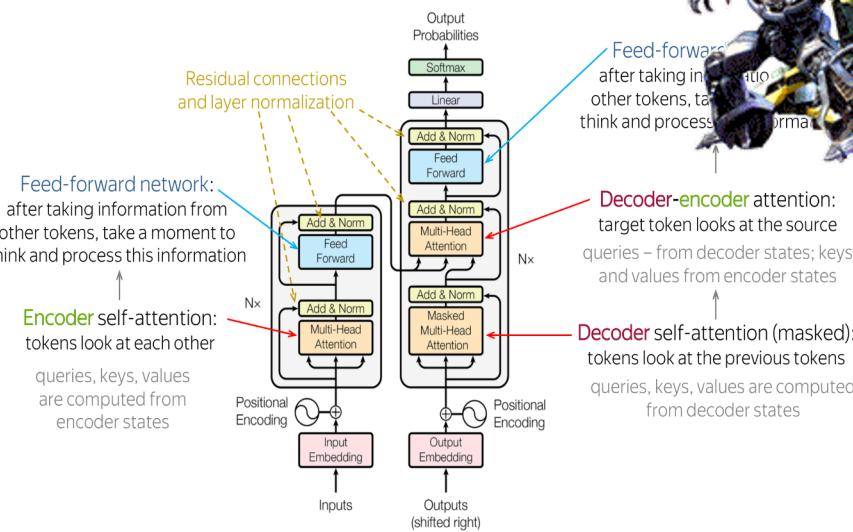
Добавляют вход блока к выходу без изменения, что упрощает течение градиентов и позволяет стековать больше слоев

## Layer Normalization для ускорения обучения:

Улучшает сходимость

# Transformer Models made revolution in NLP

- 



# Transformer LMs

The neural architecture influences the three types of pretraining, and natural use cases:

1) Encoder part

(Gets bidirectional context – can condition on future)

Spoiler: like BERT

2) Language models. Decoder part

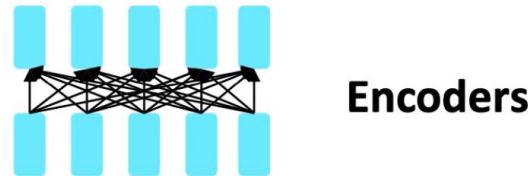
(Nice to generate from; can't condition on future words)

Spoiler: like GPTs

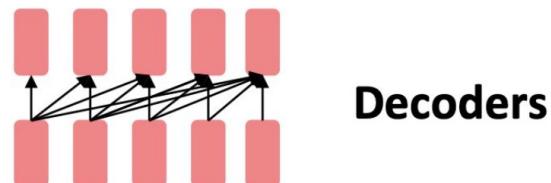
3) Encoder-decoder architecture

(best from both parts)

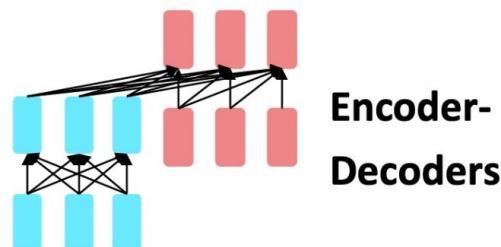
Spoiler: like T5 for example



Encoders

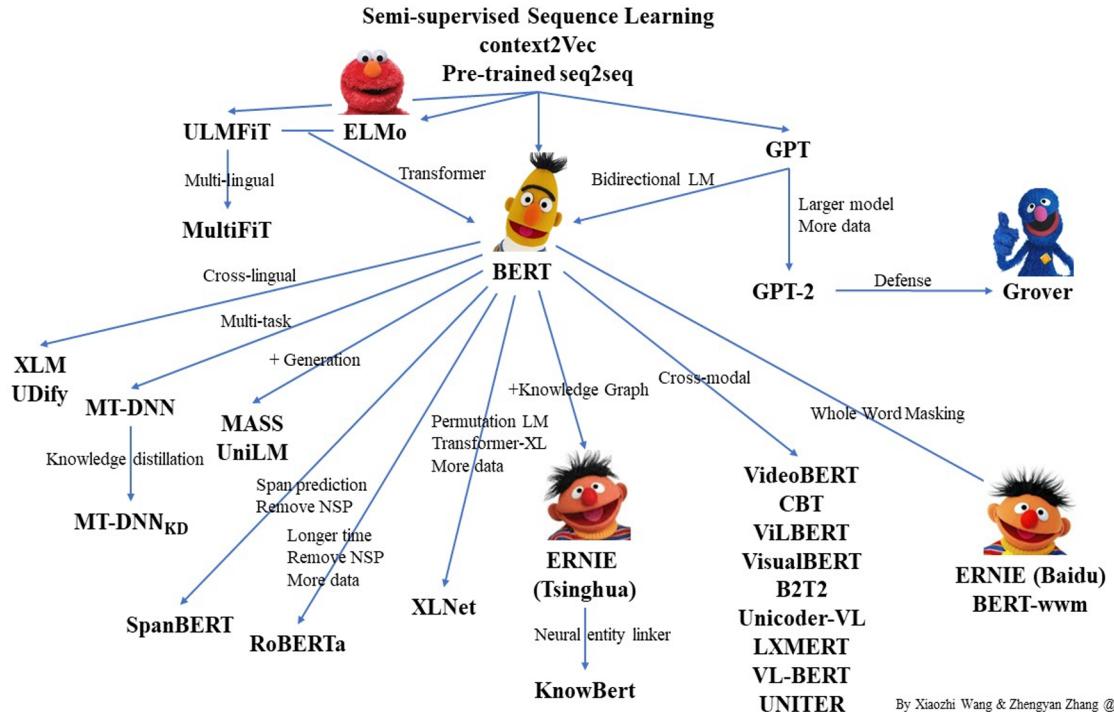


Decoders



Encoder-Decoders

# Language model Zoo

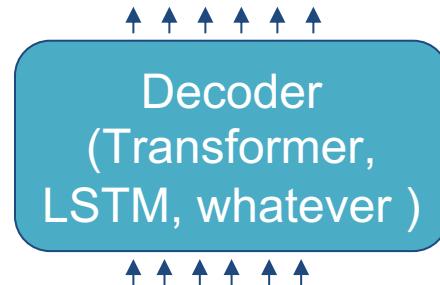


By Xiaozhi Wang & Zhengyan Zhang @THUNLP

# Pre-training/fine-tuning

- **Pre-train** through language modeling
- Train a NN to perform on a large amount of text
- Save the network parameters

*text is here for LM</eos>*

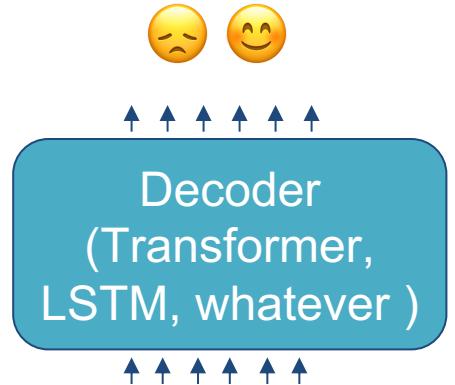


*Some text is here for LM*

# Pre-training/fine-tuning

## Fine-tune

- Take a pretrained model
- Train for the specific task (e.g., sentiment classification) with a rather small learning rate
- Make good quality using LM knowledge



*... the final of Game  
of thrones was...*

# QA systems

The goal of question answering is to build systems that automatically correctly answer questions posed by humans in a natural language.



# QA systems

- What information source does a system build on?
  - a text passage
  - all Web documents
  - knowledge bases,
  - images..
- Answer type
  - a short segment of text
  - a paragraph
  - a list
  - yes/no, ...
- Question type
  - Factoid vs non-factoid
  - open-domain vs closed-domain
  - simple vs compositional



## QA through years

<b>Setting</b>	Closed-domain	Open-domain	Reading comprehension	Open-domain	Conversational, multi-hop, multilingual
<b>Methodology</b>	Hand-engineered parsers	IR + shallow linguistic analysis	Document reader	IR + document reader	IR + document reader
<b>Systems, datasets</b>	LUNAR, QUALM	TREC QA	CNN / Daily Mail, SQuAD	Natural Questions	CoQA, TyDiQA, HotPotQA
<b>Years</b>	1970s–1990s	2000s	2013–today	2019–today	2020–today

## Processes automation and engineering

- QA goal applications
- Chatbots
- Call centers
- Study projects
- etc.

## Research and Science

- Turing test
- AI



# QA systems. Applications

Google Где находится самое глубокое озеро в мире? X Картинки Карты Новости Видео Ещё Настройки Инструменты

Результатов: примерно 498 000 (0,67 сек.)



1-е место: Байкал – это **самое глубокое озеро** России, Евразии и всего **мира**, достигающее в глубину 1642 метра. Расположенный на юге Восточной Сибири водоем является крупнейшим природным резервуаром пресной воды – он хранит в себе 20% от общего запаса поверхностной пресной воды планеты. 26 мая 2015 г.

[areal-tur.ru](#) › Италия

[Самые глубокие водоемы. Самое глубокое озеро на земле](#)

О выделенных описаниях • Оставить отзыв

[ru.wikipedia.org](#) › wiki › Список\_глубочайших\_озёр... ▾

**Список глубочайших озёр мира — Википедия**

В списке глубочайших озёр мира представлены глубочайшие озёра мира в порядке убывания их глубины. Глубочайшие озёра по частям света ...

X Ассистент

Салют! Мы – семейство виртуальных ассистентов Сбербанка. Нас здесь трое, Сбер, Джой и Афина, выбери одного из нас.

**Сбер**  
Деловой стиль общения, как у сотрудника Сбера 

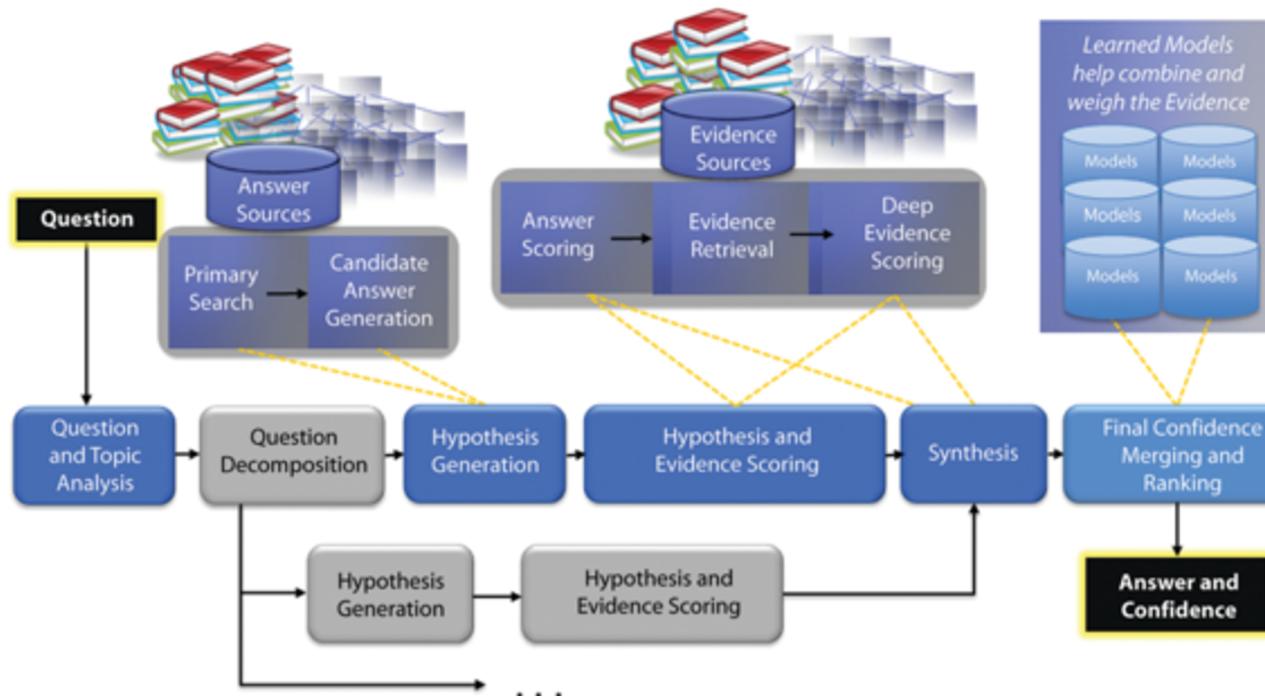
**Джой**  
Лёгкий стиль общения и бодрое настроение 

**Афина**  
Умеренный тон, понимающий собеседник для любых задач 

Здравствуйте! Вот примеры того, чем

## IBM Watson beat Jeopardy champions



# QA systems

## Question answering now



**Question:** How many parameters does BERT-large have?

**Reference Text:** BERT-large is really big... it has 24 layers and an embedding size of 1,024, for a total of 340M parameters! Altogether it is 1.34GB, so expect it to take a couple minutes to download to your Colab instance.



Almost all the state-of-the-art question answering systems are built on top of end-to-end training and pre-trained language models

# The Answer to the Ultimate Question of Life, the Universe, and Everything

42

## ANSWER

What do the answers  
look like?

## SOURCE

Where can I get the answers from?

## QUESTION

How does the question look like  
(taxonomy)?

# Answers

- Factoid
- Yes/no
- Opinion/Info
- Explanation
- Document
- A sentence or paraphraph extracted
- Another question
- etc.

# Questions

**One-hop (single-hop) question** is the question that can be answered based on a single sentence from a passage.

# Questions

**One-hop (single-hop) question** is the question that can be answered based on a single sentence from a passage.

**Multi-hop question** is a question that requires reasoning over information spread across several sentences in a passage.

*(1) Mother bought apples. (2) They were on the table. (3) John has never eaten apples, that's why he couldn't stand it and tried one.*

Question: "Where were apples that were eaten by the boy?"

# Reading comprehension

- Reading comprehension = comprehend a passage of text and answer questions about its content (P, Q) → A

"Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.  
Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum."

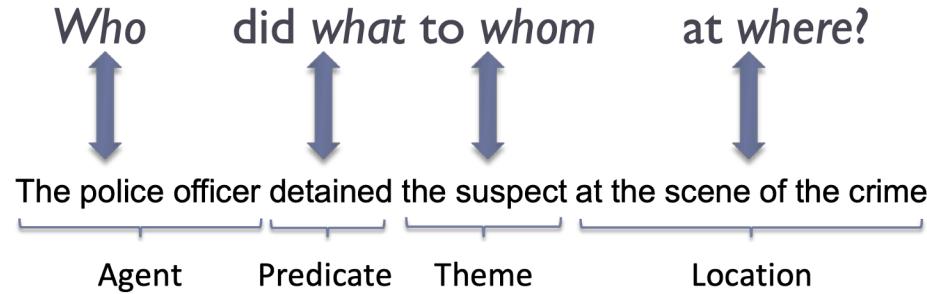
Question

- Reading comprehension task: build a system to comprehend a passage of text and answer questions about its content (P, Q) → A

# Reading comprehension

*Why Reading comprehension is important?*

- Useful in many complex practical applications
- Testbed for evaluating how well computer systems understand human language
- Many complex NLP tasks can be reduced to a reading comprehension problem:
  - Semantic Role Labeling
  - Information extraction



**Text in**

Brazil ranks number 5 in the list of countries by population.

The term "Ibu Negara" (Lady/Mother of the State) is used for wife of the President of Indonesia.

Game of Thrones is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels. It ranks fourth among the IMDB Top Rated TV Shows.

**Data out**

THE COUNTRIES WITH THE LARGEST POPULATION

China	1	1,388,232,693
India	2	1,342,512,706
United States	3	326,474,013
Indonesia	4	263,510,146
Brasil	5	174,315,386

THE COUNTRY'S FIRST LADIES

Brigitte Macron
- Spouse: Emmanuel Macron, President of France (2017 - )
Melania Trump
- Spouse: Donald J. Trump, U.S. President (2017- )
Iriana Widodo
- Spouse: Joko Widodo, President of Indonesia (2014 - )
- Also known as: "Ibu Negara" (Lady/Mother of the State)

IMDB TOP RATED TV SHOWS

- 1 Planet Earth II (2016) 9.6.
- 2 Band of Brothers (2001) 9.5.
- 3 Planet Earth (2006) 9.5.
- 4 Game of Thrones (2011) 9.4.
- 5 Breaking Bad (2008) 9.4.

# Reading comprehension

Problem formulation:

- *Input:*  $C = (c_1, c_2, \dots, c_N)$   $Q = (q_1, q_2, \dots, q_M)$   $c_i, q_i \in V$
- *Output:*  $1 \leq \text{start} \leq \text{end} \leq N$

answer is a span in the passage

- A family of LSTM-based models with attention (2016-2018)  
Attentive Reader, Stanford Attentive Reader, MatchLSTM, BiDAF, Dynamic coattention network...
- Fine-tuning Transformer models for reading comprehension (2019+)

# Reading comprehension. BERT

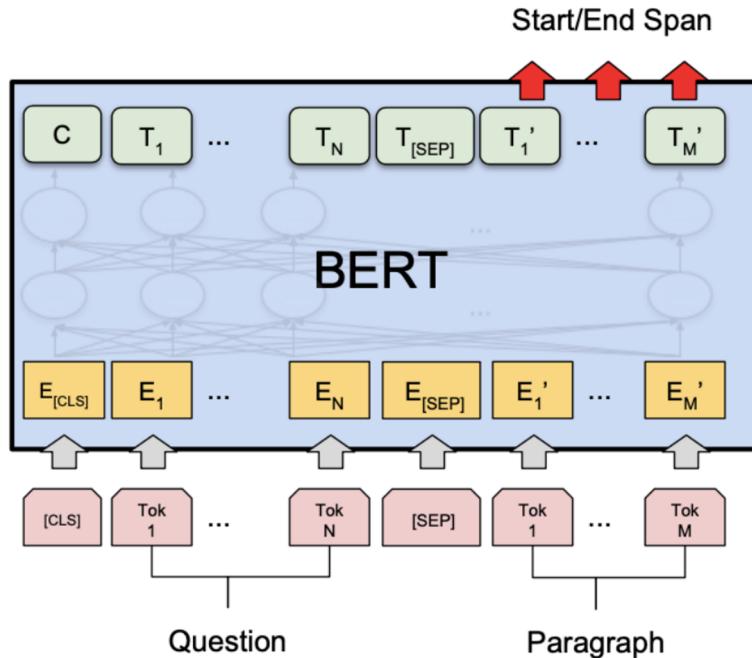
$$L = -\log p_{\text{start}}(s^*) - \log p_{\text{end}}(e^*)$$

$$p_{\text{end}}(i) = \text{softmax}_i(w^1_{\text{end}} H)$$

$$p_{\text{start}}(i) = \text{softmax}_i(w^1_{\text{start}} H)$$

where  $H = [h_1, h_2, \dots, h_N]$  are the hidden vectors of the paragraph, returned by BERT

All the BERT parameters ( $\sim 110M$ ) as well as  $H_{\text{start}}$  and  $H_{\text{end}}$  (e.g.,  $768 \times 2 = 1536$ ) are optimized together for  $L$



# Reading comprehension

## SQuAD

**Dataset size (Russian):** 50k questions

**Dataset size (English):** 100k questions

**Task:** Find the answer and direct span for the question in text

**Evaluation:** exact match (0 or 1) and F1 (partial credit)

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

## Example

**Passage:** Первая школа в Манитобе  
была основана в 1818 году  
католическими миссионерами в  
городе Виннипег, первая  
протестантская школа была  
учреждена в 1820 году.  
Провинциальное Управление  
образования было учреждено в 1871  
году, оно отвечало за  
государственные школы и учебные  
программы, ...

**Question:** Кем была в 1818 году основана первая  
школа в Манитобе?

**Answer:**

"text": "католическими миссионерами",  
"answer\_start": 50

# Reading comprehension

RuCoS

**Dataset size**

72193 train / 4370 val / 4147 test

**Data source:**

Lenta & Deutsche Welle

**Task:** Find the correct entity in the paragraph that best fits the placeholder in the query.

**Example**

**Passage:** Мать двух мальчиков, брошенных отцом в московском аэропорту Шереметьево, забрала их. Об этом сообщили TACC 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

# Reading comprehension

## MuSeRC

**Task:** Reading comprehension challenge, questions can be answered only based on multiple sentences from the paragraph.

**Dataset size:** 500/100/322

### Data source

+800 paragraphs ~6k questions

5 different domains collected from open sources:

- 1) elementary school texts
- 2) news
- 3) fiction stories
- 4) fairy tales
- 5) brief annotations of TV series and books

### 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 - немок - 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)

# Reading comprehension

## DaNetQA

**Dataset size:** 800 train, 200 dev, 200 test examples; 562 (~59%) unique questions

**Task:** Given a passage, answer a yes/no question to it.

### Data source

- 1) Crowdsourced questions are used as queries to Wikipedia
- 2) Wikipedia pages are retrieved via Google API
- 3) Passages are retrieved by Deep Pavlov SQuAD models
- 4) Crowd workers answer the questions based on the passages

### Example

**Passage:** В период с 1969 по 1972 год по программе «Аполлон» было выполнено 6 полётов с посадкой на Луне.

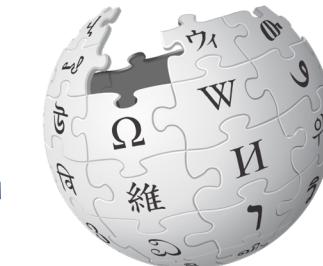
**Question:** Был ли человек на луне?

**Answer:** Yes

# Open domain question answering

- We don't assume a given passage; we have access to a large collection of documents (e.g., Wikipedia); we don't know where the answer is located
- The goal: to return the answer for any open-domain questions.
- Closed-domain
- Factoid question

# Question



WIKIPEDIA  
The Free Encyclopedia

# Google

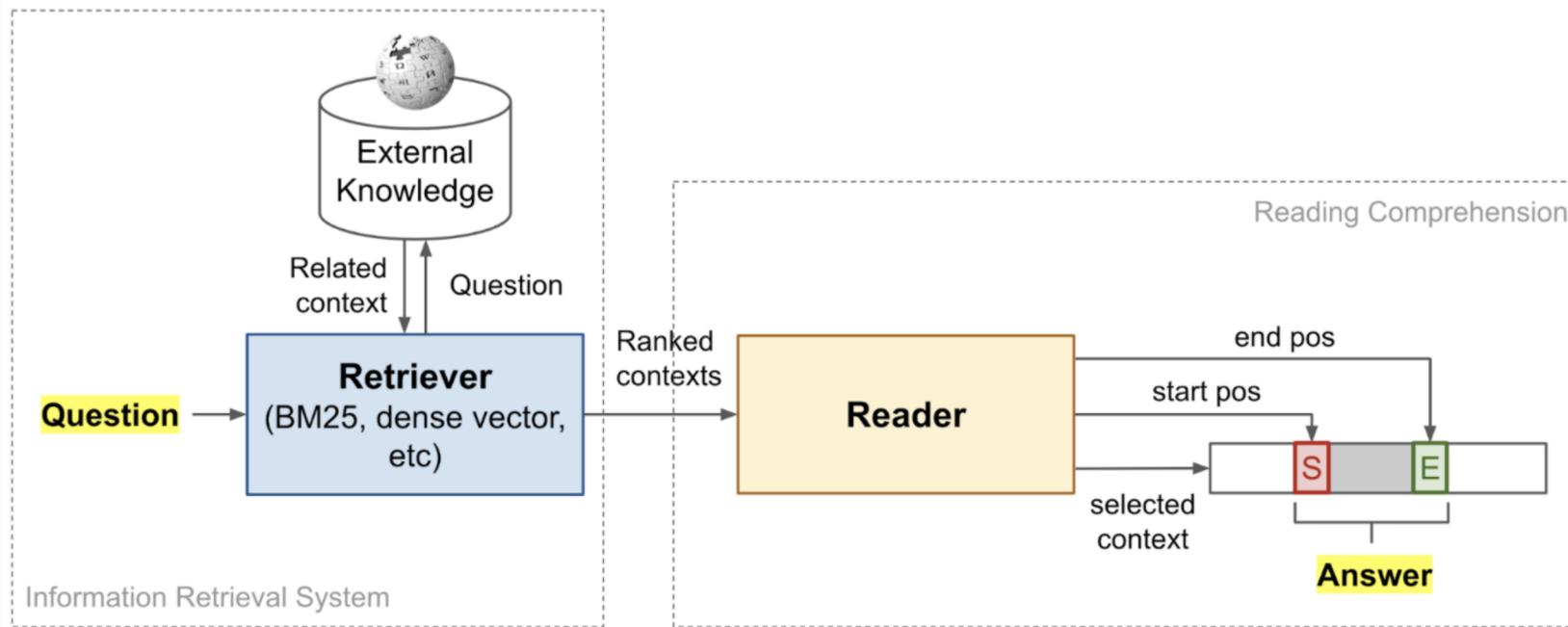
# Answer

# Open domain question answering

## Retriever-reader framework

Input: a large collection of documents  $D = D_1, D_2, \dots, D_n$  and  $Q$

Output: an answer string  $A$



# Open domain question answering

## DrQA (Document retriever Question-Answering)

Retriever: standard TF-IDF information-retrieval sparse model (a fixed module)

Reader: a neural reading comprehension model

(3-layer bidirectional LSTM with hidden size 128)

$$\text{tf-idf}(t, d, D) = \text{tf}(t, d) \times \text{idf}(t, D)$$

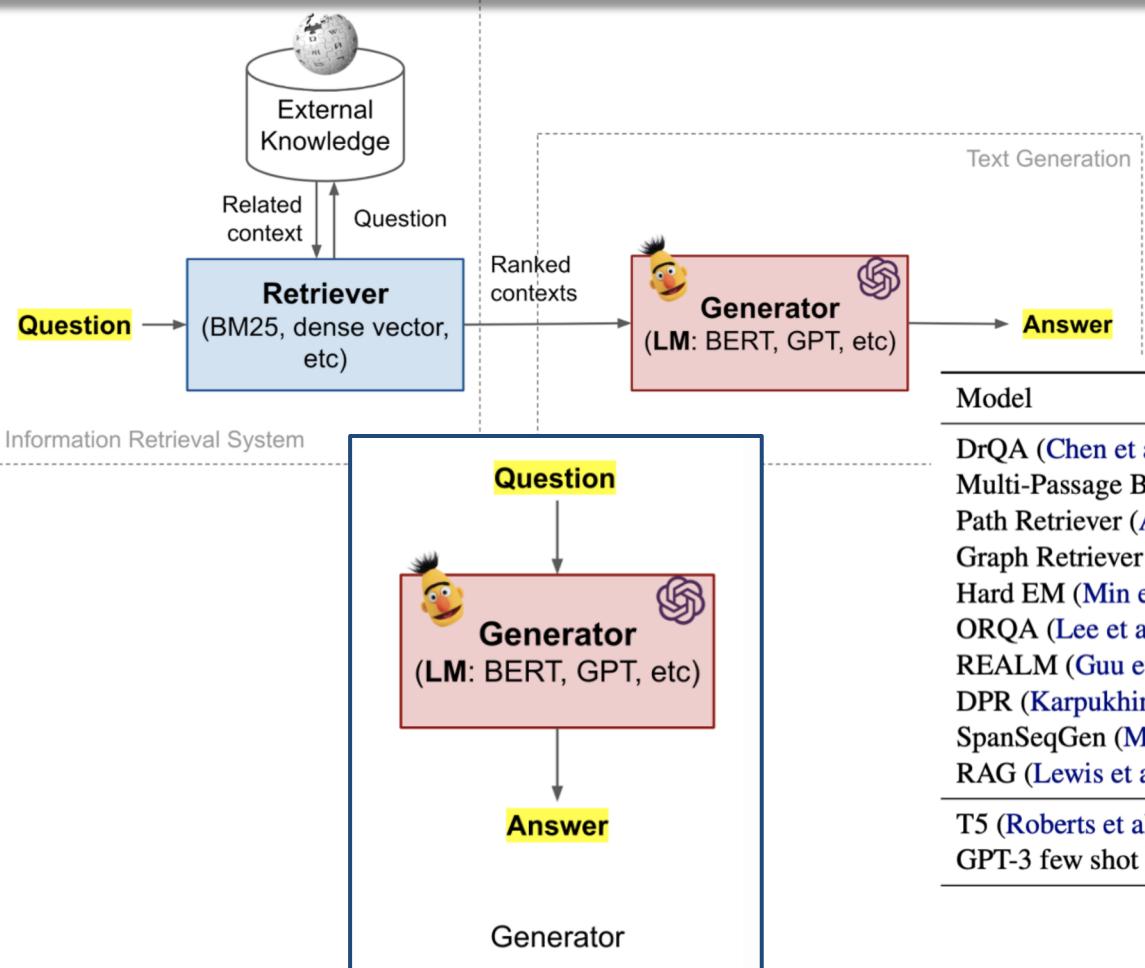
$$\text{tf}(t, d) = \log(1 + \text{freq}(t, d))$$

$$\text{idf}(t, D) = \log\left(\frac{|D|}{|\{d \in D : t \in d\}|}\right)$$

Train the retriever using question-answer pairs:

- ORQA (Open-Retrieval Question-Answering)
- DPR (Dense passage retrieval)
- REALM (Retrieval-Augmented Language Model pre-training)

# Open domain question answering



Model	NaturalQuestions	TriviaQA
DrQA (Chen et al., 2017)	-	-
Multi-Passage BERT (Wang et al., 2019)	-	-
Path Retriever (Asai et al., 2020)	31.7	-
Graph Retriever (Min et al., 2019b)	34.7	55.8
Hard EM (Min et al., 2019a)	28.8	50.9
ORQA (Lee et al., 2019)	31.3	45.1
REALM (Guu et al., 2020)	38.2	-
DPR (Karpukhin et al., 2020)	41.5	57.9
SpanSeqGen (Min et al., 2020)	42.5	-
RAG (Lewis et al., 2020)	44.5	56.1 68.0
T5 (Roberts et al., 2020)	36.6	- 60.5
GPT-3 few shot (Brown et al., 2020)	29.9	- 71.2

# Semantic Parsing

**Semantic parsing** is a process of mapping a natural language into a formal representation of its meaning. Depending of the formalism logical representation can be used to query a structured knowledge base.

Question

Logic form

Answer



# Knowledge base QA (KBQA)

KBQA - Knowledge Base question answering.

- Formal representation of knowledge.
- The graph models physical and abstract entities and relationships between them.
- A graph is defined classically as a set

$$G = (V, E) | E \subseteq \mathbb{R}^{|V| \times |V|}$$

<b>label</b>	Moscow (Q649)	item identifier																				
<b>description</b>	capital city and the largest city of Russia; separate federal subject of Russia	Moskva   Москва   Moscow, Russia   Moskva Federal City, Russia   Moscow, USSR   Moskva, Russia   City of Moscow   Moscow, Russian Federation   Moscow, Soviet Union   Moscow, Russian SFSR																				
<b>aliases</b>	<b>In more languages</b> <small>Configure</small>																					
<b>language, label, description, aliases</b>																						
<table border="1"><thead><tr><th>Language</th><th>Label</th><th>Description</th><th>Also known as</th></tr></thead><tbody><tr><td>English</td><td>Moscow</td><td>capital city and the largest city of Russia; separate federal subject of Russia</td><td>Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Москва, Россия City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR</td></tr><tr><td>Russian</td><td>Москва</td><td>столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)</td><td>Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим</td></tr><tr><td>German</td><td>Moskau</td><td>Hauptstadt von Russland</td><td></td></tr><tr><td>French</td><td>Moscou</td><td>capitale de la Russie</td><td></td></tr></tbody></table>			Language	Label	Description	Also known as	English	Moscow	capital city and the largest city of Russia; separate federal subject of Russia	Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Москва, Россия City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR	Russian	Москва	столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)	Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим	German	Moskau	Hauptstadt von Russland		French	Moscou	capitale de la Russie	
Language	Label	Description	Also known as																			
English	Moscow	capital city and the largest city of Russia; separate federal subject of Russia	Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Москва, Россия City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR																			
Russian	Москва	столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)	Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим																			
German	Moskau	Hauptstadt von Russland																				
French	Moscou	capitale de la Russie																				
<a href="#">All entered languages</a>																						

# Knowledge base QA (KBQA)

**KBQA** - Knowledge Base question answering.

- Formal representation of knowledge.
- The graph models physical and abstract entities and relationships between them.
- A graph is defined classically as a set

$$G = (V, E) | E \subseteq \mathbb{R}^{|V| \times |V|}$$

For example, Wikidata - graph db, cross-links in Wikipedia.

DBpedia, Wikidata, YAGO, etc.

<b>label</b>	Moscow (Q649)	item identifier												
<b>description</b>	capital city and the largest city of Russia; separate federal subject of Russia													
<b>aliases</b>	Moskva   Москва   Moscow, Russia   Moskva Federal City, Russia   Moscow, USSR   Moskva, Russia   City of Moscow   Moscow, Russian Federation   Moscow, Soviet Union   Moscow, Russian SFSR													
<b>In more languages</b> <small>Configure</small>														
<table border="1"><thead><tr><th>Language</th><th>Label</th><th>Description</th><th>Also known as</th></tr></thead><tbody><tr><td>English</td><td>Moscow</td><td>capital city and the largest city of Russia; separate federal subject of Russia</td><td>Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Moskva, Russia City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR</td></tr></tbody></table>			Language	Label	Description	Also known as	English	Moscow	capital city and the largest city of Russia; separate federal subject of Russia	Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Moskva, Russia City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR				
Language	Label	Description	Also known as											
English	Moscow	capital city and the largest city of Russia; separate federal subject of Russia	Moskva Москва Moscow, Russia Moskva Federal City, Russia Moscow, USSR Moskva, Russia City of Moscow Moscow, Russian Federation Moscow, Soviet Union Moscow, Russian SFSR											
<b>language, label, description, aliases</b>														
<table border="1"><tbody><tr><td>Russian</td><td>Москва</td><td>столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)</td><td>Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим</td></tr><tr><td>German</td><td>Moskau</td><td>Hauptstadt von Russland</td><td></td></tr><tr><td>French</td><td>Moscou</td><td>capitale de la Russie</td><td></td></tr></tbody></table>			Russian	Москва	столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)	Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим	German	Moskau	Hauptstadt von Russland		French	Moscou	capitale de la Russie	
Russian	Москва	столица и крупнейший город России; город федерального значения; административный центр Московской области (не входит в её состав)	Первопрестольная Порт пяти морей Москва (город) Москва, Россия Москва (Россия) Москва Златоглавая Третий Рим											
German	Moskau	Hauptstadt von Russland												
French	Moscou	capitale de la Russie												
<a href="#">All entered languages</a>														

# Knowledge base QA

The Resource Description Framework (**RDF**) is a model that defines the model of the *subject-predicate-subject* or *subject-predicate-object* triplet.

That is, an entity - "subject" can be associated with another entity or a simple value - an object - through some property - a predicate.

# Knowledge base QA

The Resource Description Framework (**RDF**) is a model that defines the model of the *subject-predicate-subject* or *subject-predicate-object triplet*.

That is, an entity - "subject" can be associated with another entity or a simple value - an object - through some property - a predicate.

Special predicates: *rdf:type*, *rdf:Property*, *rdf:subject*, *rdf:predicate*, *rdf:object*, *rdf:first*, *rdf:value*, *rdf>List*, etc..

Triplet example:

"Университет ИТМО - находится в - Санкт-Петербург" links entities: Университет ИТМО and Санкт-Петербург via predicate "находится в".

*Triplet "Университет ИТМО - rdf:type - Университет"* means that "Университет ИТМО" ∈ университеты.

# Knowledge base QA

The **Resource Description Framework (RDF)** is a model that defines the model of the *subject-predicate-subject* or *subject-predicate-object* triplet.

That is, an entity - "subject" can be associated with another entity or a simple value - an object - through some property - a predicate.

Special predicates: *rdf:type*, *rdf:Property*, *rdf:subject*, *rdf:predicate*, *rdf:object*, *rdf:first*, *rdf:value*, *rdf>List*, etc..

Triplet example:

"Университет ИТМО - находится в - Санкт-Петербург" links entities: Университет ИТМО and Санкт-Петербург via predicate "находится в".

Triplet "Университет ИТМО - *rdf:type* - Университет" means that "Университет ИТМО" ∈ университеты.

To query the knowledge represented in RDF, the query language **SPARQL** is used (links knowledge graphs to applications based on knowledge graphs)

The part of the knowledge graph that describes abstract concepts and connections between them at a high level, otherwise it is also called **ontology**.

# Chatbots and dialogue systems. Types

Dialogue systems, or conversational agents communicate with users in natural language (text, speech, or both)

Two classes:

**Task-oriented dialogue agents** use conversation with users to help people complete tasks.

- give directions
- control appliances
- find restaurants
- make calls etc.

**Chatbots** are systems for conversations, which mimic everyday conversations or ‘chats’

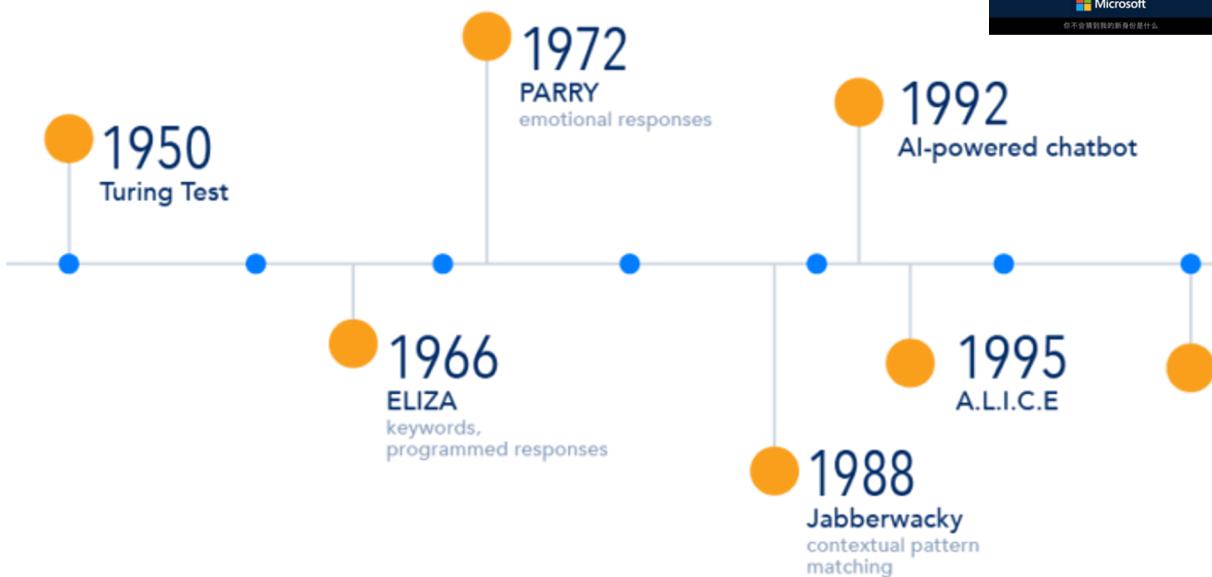
# Chatbots. Evolution

## From rule-based to NN

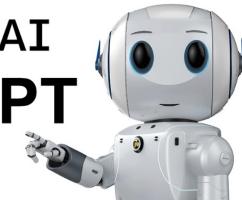


# Chatbots. Evolution

## From rule-based to NN

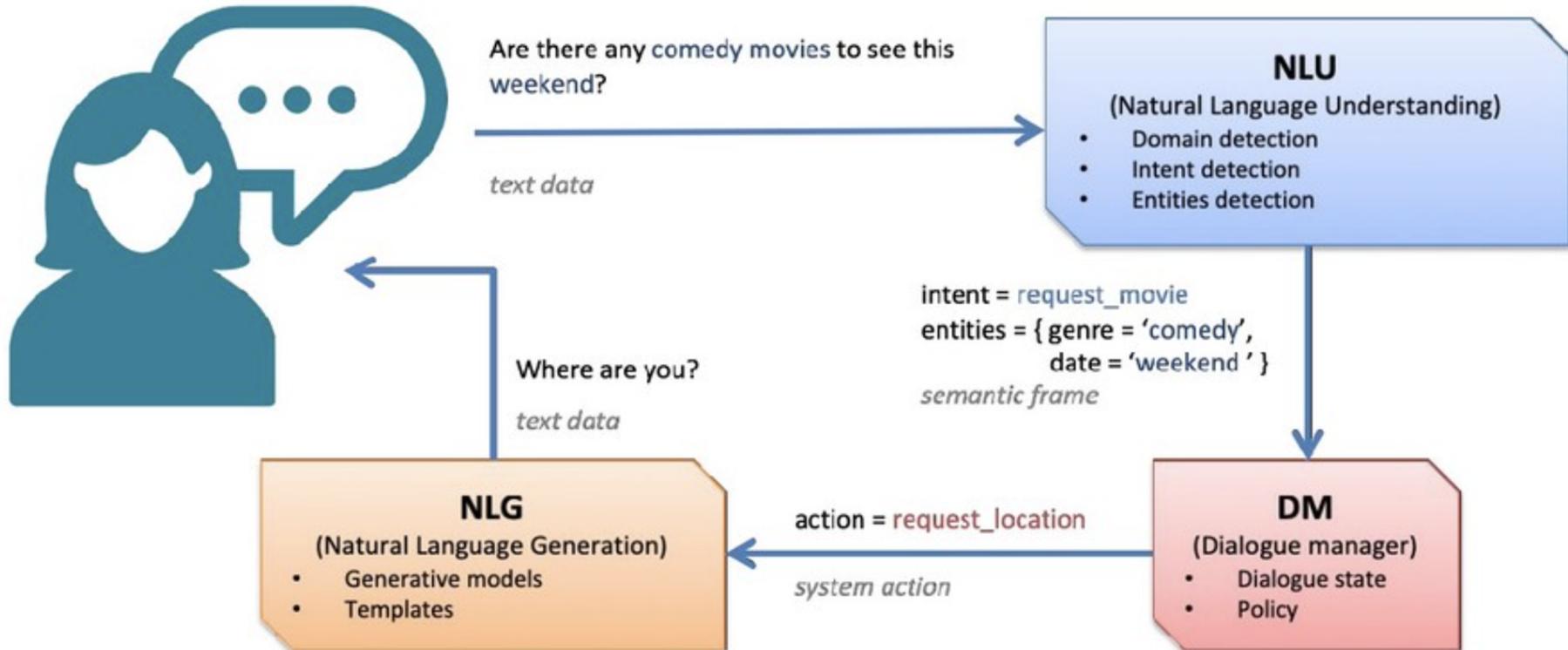


OpenAI  
**ChatGPT**



# Chatbots. Architecture

## Goal oriented chatbots



# Chatbots. Intents

**Intent recognition** or **intent** classification — is the task of taking a written or spoken input, and classifying it based on what the user wants to achieve.

**Intent recognition** forms an essential component of chatbots and finds use in sales conversions, customer support, and many other areas.

We need to do **slot filling**: extract the particular slots and fillers that the user intends the system to understand from their utterance with respect to their intent

Datasets: **SNIPS**, **ATIS**, [\*\*MultiWoZ\*\*](#)

*Show me morning flights from Moscow to Barcelona on Tuesday*

DOMAIN:	TRAVEL
INTENT:	SHOW-FLIGHT (want to travel)
DATE:	Tuesday
CITY-from:	Moscow
CITY-to:	Barcelona

# Chatbots. Anaphora

**Coreference/Anaphora Resolution** is the problem of resolving what a pronoun, or a noun phrase refers to.

Examples:

*Ann saw Mary yesterday. She was tired.*

*Mary went home. She did so because she was tired.*

*Putin made a surprise visit to USA yesterday. President met ...*

**Elipsis** is a case of anaphora., where the anaphor is a null proform.

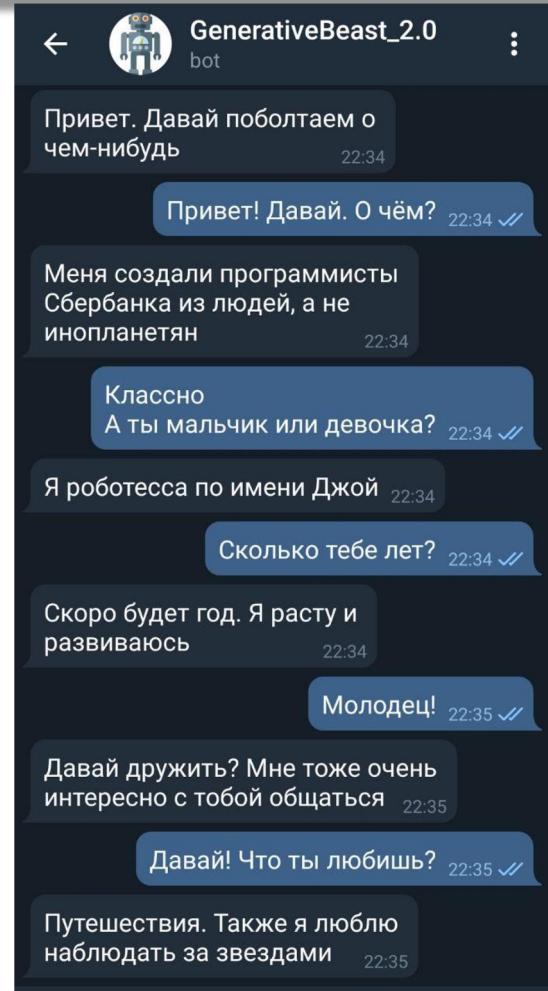
Examples:

*I haven't seen John yet, but I will Ø (see) soon.*

*Robert cooked the first course, and Mary Ø (cooked) the dessert.*

# Chatbots. Chit-chat

- Retrieval base models:
  - *Take a dataset of dialogues,*
  - *construct pairs (context, reply)*
  - *build inverted index on contexts*
  - *return best with cos\_sim(user\_context, index\_context)*
- Neural networks
- Generation based networks (GPTs)



# ChatGPT

 Расскажи анекдот про машинное обучение

 Есть такой анекдот: приходит парень к психологу и говорит:

— Доктор, у меня проблема! Я не могу перестать думать о машинном обучении! Мне даже снится как я обучаю модели и анализирую данные!

Психолог говорит:

— Не беспокойтесь, это нормально. Ведь вы робот.

— Не беспокойтесь это нормально ведь вы робот.

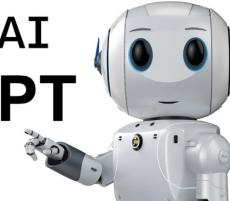
 Каковы основные пути развития искусственного интеллекта в области медицины?

 Существует несколько путей развития искусственного интеллекта в области медицины, включая:

1. Использование машинного обучения для анализа медицинских данных, таких как изображения, лабораторные тесты и электрокардиограммы, чтобы обнаруживать болезни и подбирать наиболее эффективные способы лечения.
2. Использование навыков естественного языка для создания чат-ботов и виртуальных медицинских помощников, которые могут отвечать на вопросы пациента в любое время суток.

Некоторые примеры использования AI в медицине включают в себя автоматизированную диагностику, подбор терапии на основе данных о пациенте, а также помощь врачам в принятии решений.

 OpenAI  
ChatGPT



 Придумай слоган для компании, занимающейся выпуском бензопил для сотрудников ИТ компаний

 "Бензопила от нашей компании – ваши сотрудники пилят код в два раза быстрее!"

 Examples by @doomgrad <= subscribe  
<https://openai.com/blog/chatgpt/>

# Questions?



@mashkka\_ds <= subscribe

# References. Datasets

- SQuAD 2.0: the Stanford QA dataset.
- RACE: a reading comprehension dataset collected from English Examinations that are created for middle school and high school students.
- TREC QA: the TREC QA collections.
- MS MARCO: a QA dataset featuring 100,000 real Bing questions and a human generated answer.
- CuratedTREC: based on the benchmarks from the TREC QA tasks that have been curated by Baudis & Sedivy (2015).
- Google Natural Questions: contains real user questions issued to Google search, and answers found from Wikipedia by annotators.
- WebQuestions: designed for knowledge-base QA with answers restricted to Freebase entities.
- WikiQA: Bing query logs were used as the source of questions. Each question is then linked to a Wikipedia page that potentially contains the answer.
- WikiMovies: contains movie-related questions from the OMDb and MovieLens databases and where the questions can be answered using Wikipedia pages.
- WikiReading: to predict textual values from the structured knowledge base Wikidata by reading the text of the corresponding Wikipedia articles.
- TriviaQA: a reading comprehension dataset containing 95K question-answer pairs authored by trivia enthusiasts and independently gathered multiple evidence documents per question.
- Jeopardy! Questions: contains 200,000+ Jeopardy! questions.
- DeepMind Q&A Dataset: question/answer pairs from CNN and Daily Mail articles.
- bAbi: a rich collection of datasets for text understanding by Facebook.
- FEVER: for fact extraction and verification.
- SearchQA: question-answer pairs were crawled from J! Archive, and then augmented with text snippets from Google.
- Quasar-T: a collection of open-domain trivia questions and their answers obtained from various internet sources.
- Quiz bowl: contains data from a trivia competition called quiz bowl.
- AmbigNQ: ambiguous questions selected from NQ-OPEN dataset.
- QA-Overlap: a collection of overlapped answers/questions between train and test set for Natural Questions, TriviaQA, and WebQuestions
- VQA : visual question answering. Stanford.
- MultiWOZ (The Multi-domain Wizard-of-Oz (MultiWOZ)) - <https://paperswithcode.com/dataset/multiwoz>
- RUSSIAN datasets:
  - a. SberSQuAD - <https://drive.google.com/drive/u/1/folders/1AtLPhazqhpHTC-be10XsYIKE3n1Xut51>
  - b. RuCoS - [https://russiansuperglue.com/tasks/task\\_info/RuCoS](https://russiansuperglue.com/tasks/task_info/RuCoS)
  - c. MuSeRC - [https://russiansuperglue.com/tasks/task\\_info/MuSeRC](https://russiansuperglue.com/tasks/task_info/MuSeRC)
  - d. DaNetQA - [https://russiansuperglue.com/tasks/task\\_info/DaNetQA](https://russiansuperglue.com/tasks/task_info/DaNetQA)
  - e. RuBQ - <https://github.com/vladislavneon/RuBQ>

# References. Articles

- [1] Danqi Chen & Scott Yih. "[ACL2020 Tutorial: Open-Domain Question Answering](#)" July 2020.
- [3] Shuohang Wang, et al. "[R^3: Reinforced Ranker-Reader for Open-Domain Question Answering](#)" AAAI 2018.
- [4] Jimmy Lin. "[The neural hype and comparisons against weak baselines.](#)" ACM SIGIR Forum. Vol. 52. No. 2. 2019.
- [5] Wei Yang, et al. "[End-to-End Open-Domain Question Answering with BERTserini](#)" NAACL 2019.
- [6] Christopher Clark & Matt Gardner. "[Simple and Effective Multi-Paragraph Reading Comprehension.](#)" arXiv:1710.10723 (2017).
- [7] Rodrigo Nogueira & Kyunghyun Cho. "[Passage Re-ranking with BERT.](#)" arXiv preprint arXiv:1901.04085 (2019). | [code](#)
- [8] Zhiguo Wang, et al. "[Multi-passage BERT: A globally normalized BERT model for open-domain question answering.](#)" EMNLP 2019.
- [9] Fenogenova, Alena, Vladislav Mikhailov, and Denis Shevelev. "Read and Reason with MuSeRC and RuCoS: Datasets for Machine Reading Comprehension for Russian." *Proceedings of the 28th International Conference on Computational Linguistics*. 2020.
- [10] Seo, Minjoon, et al. "Bidirectional attention flow for machine comprehension." *arXiv preprint arXiv:1611.01603* (2016).

# References. Sources

Wikidata

Wikidata Query Service

Python libraries qwikidata

Knowledge graphs course <https://ods.ai/tracks/kgcourse2021/>

DeepPavlov demos

IBM Watson video in Jeopardy challenge

Xiaoice chatbot <https://arxiv.org/abs/1812.08989>

DialoGPT2 <https://github.com/vlarine/ruDialoGPT>

Anaphora resolution for Russian (<http://www.dialog-21.ru/en/evaluation/2019/disambiguation/anaphora/>, AGRR-2019, RuCor)

Slot fillings, intent recognition [http://nlpprogress.com/english/intent\\_detection\\_slot\\_filling.html](http://nlpprogress.com/english/intent_detection_slot_filling.html)