Question Answering: From the basics to the state-of-the-art with PrimeQA

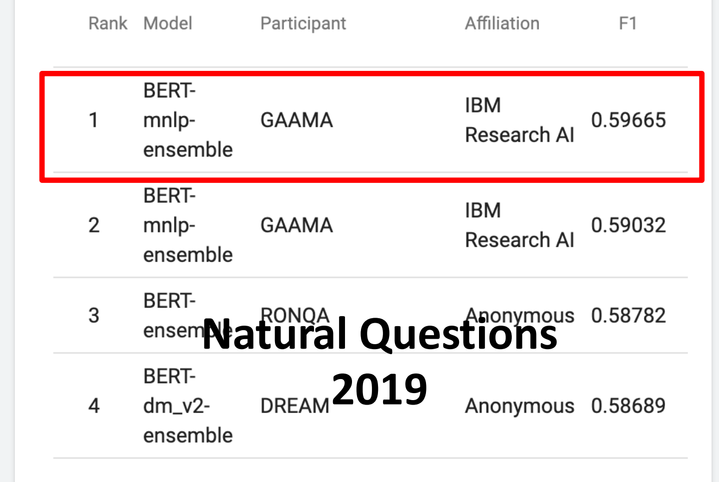
By: Avi Sil

Principal Research Scientist & Manager

https://github.com/primeqa/primeqa

Question Answering milestones at IBM Research AI

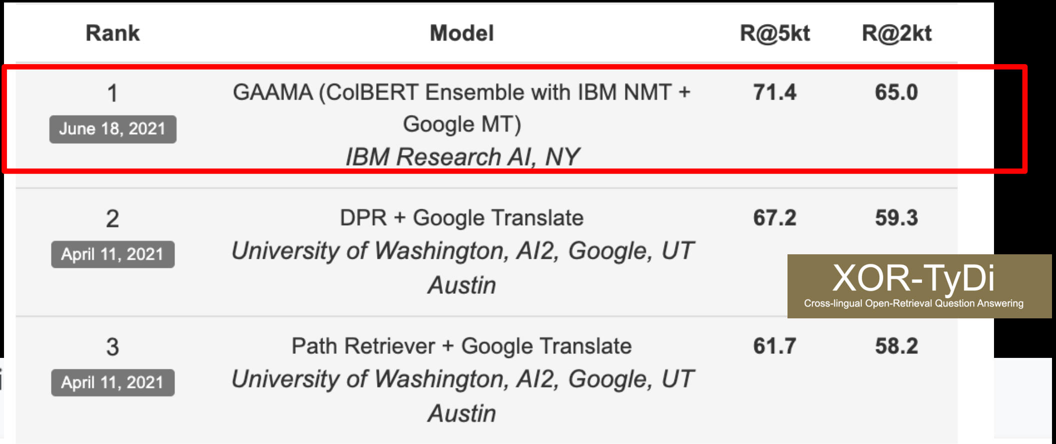
2011



Rank	Model	Participant	Affiliation	F1
1	BERT-mnlp-ensemble	GAAMA	IBM Research AI	0.59665
2	BERT-mnlp-ensemble	GAAMA	IBM Research AI	0.59032
3	BERT-ensem	RONQA	Anonymous	0.58782
4	BERT-dm_v2-ensemble	DREAM	Anonymous	0.58689

Natural Questions
2019

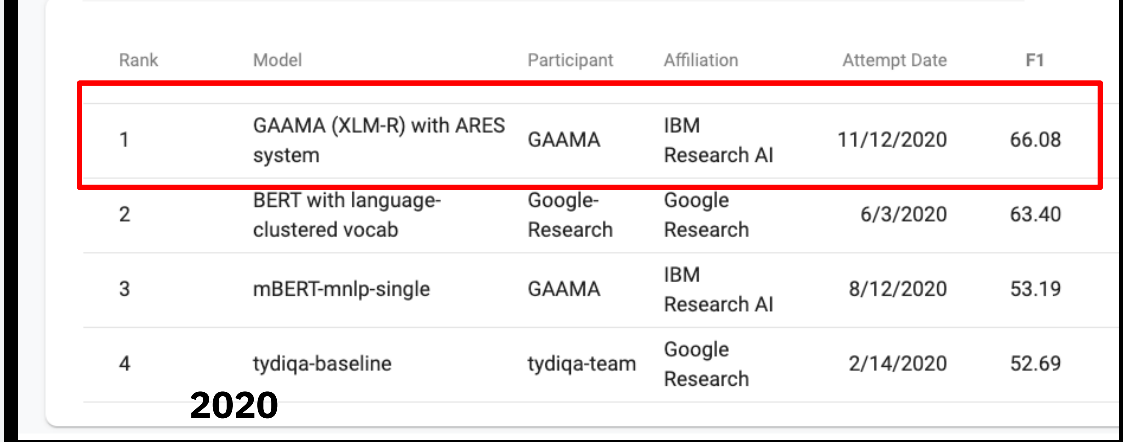
2021-2022



Rank	Model	R@5kt	R@2kt
1 June 18, 2021	GAAMA (ColBERT Ensemble with IBM NMT + Google MT) *IBM Research AI, NY*	71.4	65.0
2 April 11, 2021	DPR + Google Translate *University of Washington, AI2, Google, UT Austin*	67.2	59.3
3 April 11, 2021	Path Retriever + Google Translate *University of Washington, AI2, Google, UT Austin*	61.7	58.2

XOR-TyDi
Cross-lingual Open-Retrieval Question Answering

Mi



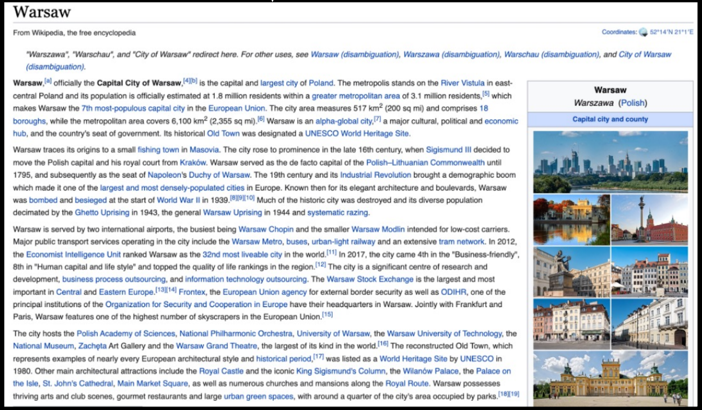
Rank	Model	Participant	Affiliation	Attempt Date	F1
1	GAAMA (XLM-R) with ARES system	GAAMA	IBM Research AI	11/12/2020	66.08
2	BERT with language-clustered vocab	Google-Research	Google Research	6/3/2020	63.40
3	mBERT-mnlp-single	GAAMA	IBM Research AI	8/12/2020	53.19
4	tydiqa-baseline	tydiqa-team	Google Research	2/14/2020	52.69

2020

What is QA? Reading Comprehension vs Open-Retrieval QA

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

Document Retriever



Warsaw

From Wikipedia, the free encyclopedia

Coordinates: 52°14′N 21°1′E

"Warszawa", "Warschau", and "City of Warsaw" redirect here. For other uses, see Warsaw (disambiguation), Warszawa (disambiguation), Warschau (disambiguation), and City of Warsaw (disambiguation).

Warsaw,[a] officially the **Capital City of Warsaw**,[4][b] is the capital and largest city of Poland. The metropolis stands on the River Vistula in east-central Poland and its population is officially estimated at 1.8 million residents within a greater metropolitan area of 3.1 million residents,[5] which makes Warsaw the 7th most-populous capital city in the European Union. The city area measures 517 km^2 (200 sq mi) and comprises 18 boroughs, while the metropolitan area covers 6,100 km^2 (2,355 sq mi).[6] Warsaw is an alpha-global city,[7] a major cultural, political and economic hub, and the country's seat of government. Its historical Old Town was designated a UNESCO World Heritage Site.

Warsaw traces its origins to a small fishing town in Masovia. The city rose to prominence in the late 16th century, when Sigismund III decided to move the Polish capital and his royal court from Kraków. Warsaw served as the de facto capital of the Polish–Lithuanian Commonwealth until 1795, and subsequently as the seat of Napoleon's Duchy of Warsaw. The 19th century and its Industrial Revolution brought a demographic boom which made it one of the largest and most densely-populated cities in Europe. Known then for its elegant architecture and boulevards, Warsaw was bombed and besieged at the start of World War II in 1939.[8][9][10] Much of the historic city was destroyed and its diverse population decimated by the Ghetto Uprising in 1943, the general Warsaw Uprising in 1944 and systematic razing.

Warsaw is served by two international airports, the busiest being Warsaw Chopin and the smaller Warsaw Modlin intended for low-cost carriers. Major public transport services operating in the city include the Warsaw Metro, buses, urban-light railway and an extensive tram network. In 2012, the Economist Intelligence Unit ranked Warsaw as the 32nd most liveable city in the world.[11] In 2017, the city came 4th in the "Business-friendly", 8th in "Human capital and life style" and topped the quality of life rankings in the region.[12] The city is a significant centre of research and development, business process outsourcing, and information technology outsourcing. The Warsaw Stock Exchange is the largest and most important in Central and Eastern Europe.[13][14] Frontex, the European Union agency for external border security as well as ODIHR, one of the principal institutions of the Organization for Security and Cooperation in Europe have their headquarters in Warsaw. Jointly with Frankfurt and Paris, Warsaw features one of the highest number of skyscrapers in the European Union.[15]

The city hosts the Polish Academy of Sciences, National Philharmonic Orchestra, University of Warsaw, the Warsaw University of Technology, the National Museum, Zachęta Art Gallery and the Warsaw Grand Theatre, the largest of its kind in the world.[16] The reconstructed Old Town, which represents examples of nearly every European architectural style and historical period,[17] was listed as a World Heritage Site by UNESCO in 1980. Other main architectural attractions include the Royal Castle and the iconic King Sigismund's Column, the Wilanów Palace, the Palace on the Isle, St. John's Cathedral, Main Market Square, as well as numerous churches and mansions along the Royal Route. Warsaw possesses thriving arts and club scenes, gourmet restaurants and large urban green spaces, with around a quarter of the city's area occupied by parks.[18][19]

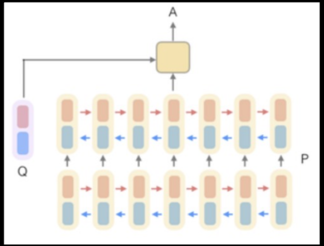
Warsaw

Warszawa (Polish)

Capital city and county

Document Reader

833,500



Open-Retrieval QA (ORQA)

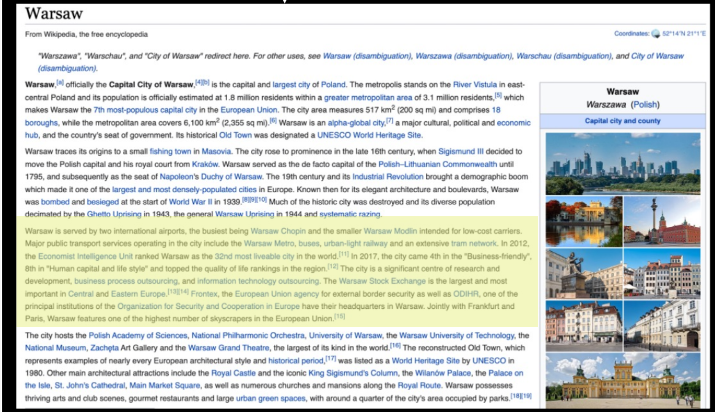
Note: ORQA is aka Open Domain QA [Lee et al., 2019] and/or End-2-end QA [Reddy et al., 2021].

What is QA? Reading Comprehension vs Open-Retrieval QA

Document Retriever

Reading comprehension

Q: How many of Warsaw's inhabitants spoke Polish in 1933?



Warsaw

From Wikipedia, the free encyclopedia

Coordinates: 52°14′N 21°1′E

"Warszawa", "Warschau", and "City of Warsaw" redirect here. For other uses, see Warsaw (disambiguation), Warszawa (disambiguation), Warschau (disambiguation), and City of Warsaw (disambiguation).

Warsaw,[a] officially the **Capital City of Warsaw**,[4][b] is the capital and largest city of Poland. The metropolis stands on the River Vistula in east-central Poland and its population is officially estimated at 1.8 million residents within a greater metropolitan area of 3.1 million residents,[5] which makes Warsaw the 7th most-populous capital city in the European Union. The city area measures 517 km^2 (200 sq mi) and comprises 18 boroughs, while the metropolitan area covers 6,100 km^2 (2,355 sq mi).[6] Warsaw is an alpha-global city,[7] a major cultural, political and economic hub, and the country's seat of government. Its historical Old Town was designated a UNESCO World Heritage Site.

Warsaw traces its origins to a small fishing town in Masovia. The city rose to prominence in the late 16th century, when Sigismund III decided to move the Polish capital and his royal court from Kraków. Warsaw served as the de facto capital of the Polish–Lithuanian Commonwealth until 1795, and subsequently as the seat of Napoleon's Duchy of Warsaw. The 19th century and its Industrial Revolution brought a demographic boom which made it one of the largest and most densely-populated cities in Europe. Known then for its elegant architecture and boulevards, Warsaw was bombed and besieged at the start of World War II in 1939.[8][9][10] Much of the historic city was destroyed and its diverse population decimated by the Ghetto Uprising in 1943, the general Warsaw Uprising in 1944 and systematic razing.

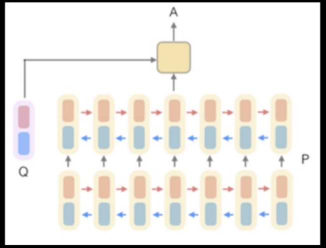
Warsaw is served by two international airports, the busiest being Warsaw Chopin and the smaller Warsaw Modlin intended for low-cost carriers. Major public transport services operating in the city include the Warsaw Metro, buses, urban-light railway and an extensive tram network. In 2012, the Economist Intelligence Unit ranked Warsaw as the 32nd most liveable city in the world.[11] In 2017, the city came 4th in the "Business-friendly", 8th in "Human capital and life style" and topped the quality of life rankings in the region.[12] The city is a significant centre of research and development, business process outsourcing, and information technology outsourcing. The Warsaw Stock Exchange is the largest and most important in Central and Eastern Europe.[13][14] Frontex, the European Union agency for external border security as well as ODIHR, one of the principal institutions of the Organization for Security and Cooperation in Europe have their headquarters in Warsaw. Jointly with Frankfurt and Paris, Warsaw features one of the highest number of skyscrapers in the European Union.[15]

The city hosts the Polish Academy of Sciences, National Philharmonic Orchestra, University of Warsaw, the Warsaw University of Technology, the National Museum, Zachęta Art Gallery and the Warsaw Grand Theatre, the largest of its kind in the world.[16] The reconstructed Old Town, which represents examples of nearly every European architectural style and historical period,[17] was listed as a World Heritage Site by UNESCO in 1980. Other main architectural attractions include the Royal Castle and the iconic King Sigismund's Column, the Wilanów Palace, the Palace on the Isle, St. John's Cathedral, Main Market Square, as well as numerous churches and mansions along the Royal Route. Warsaw possesses thriving arts and club scenes, gourmet restaurants and large urban green spaces, with around a quarter of the city's area occupied by parks.[18][19]

Warsaw
Warszawa (Polish)
Capital city and county

Document Reader

833,500



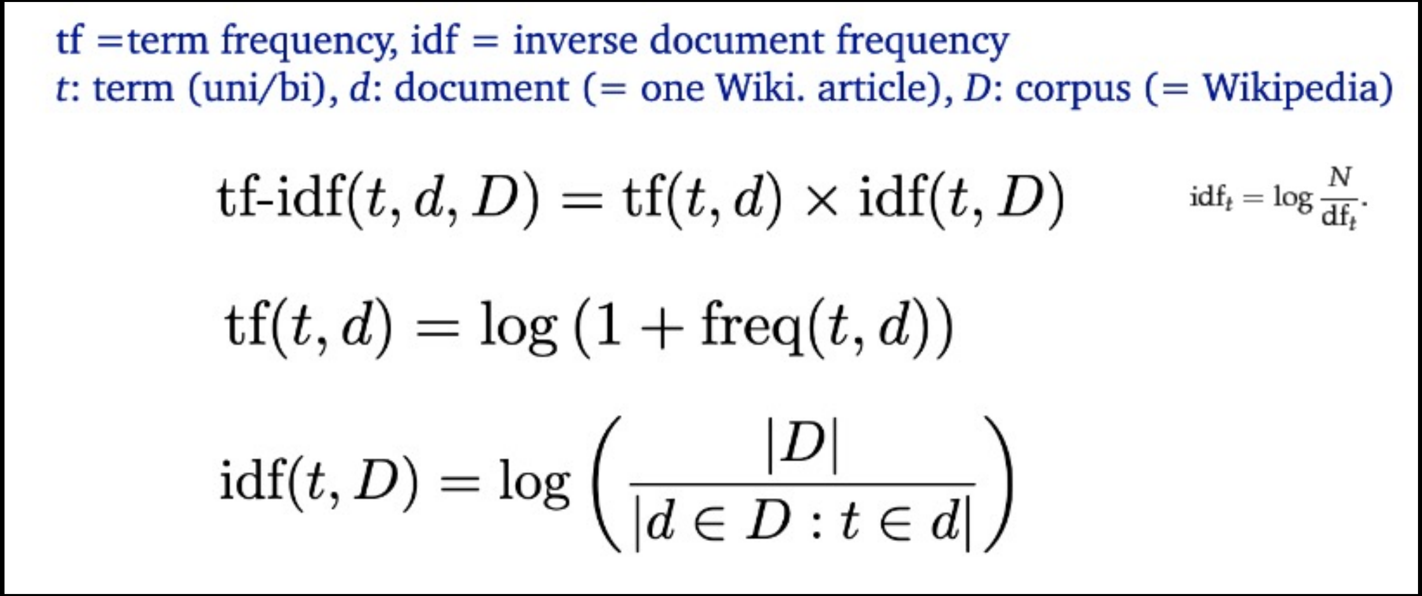
Link to IBM Research's GAAMA (public) demos

- GAAMA : Go Ahead Ask Me Anything
    - Reading Comprehension (English only): http://ibm.biz/ibm_gaama
    - ORQA Covid-19 Demo: http://ibm.biz/covidAnswerFinding

Retrievers – Nuts and bolts

A Traditional Retriever

- A TF-IDF [Robertson 2004] weighted term vector model over unigrams/ bi-grams



tf =term frequency, idf = inverse document frequency
t: term (uni/bi), d: document (= one Wiki. article), D: corpus (= Wikipedia)

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

$$\text{idf}_t = \log \frac{N}{\text{df}_t}.$$

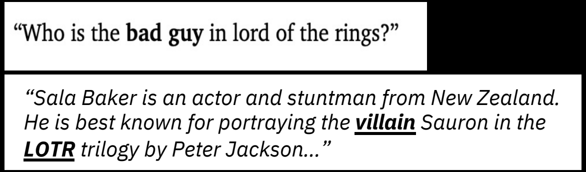
$$\text{tf}(t, d) = \log(1 + \text{freq}(t, d))$$

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

- However, this retriever is not trainable

However, they have limitations!

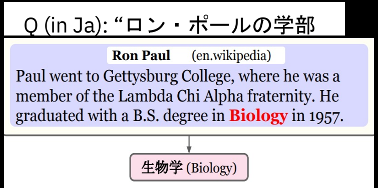
1. Can NOT answer questions when there’s little or no lexical overlap



“Who is the **bad guy** in lord of the rings?”

*“Sala Baker is an actor and stuntman from New Zealand. He is best known for portraying the **villain** Sauron in the **LOTR** trilogy by Peter Jackson...”*

2. Can NOT retrieve **cross-lingual** passages without translation (needs special models)



Q (in Ja): “ロン・ポールの学部

Ron Paul (en.wikipedia)

Paul went to Gettysburg College, where he was a member of the Lambda Chi Alpha fraternity. He graduated with a B.S. degree in **Biology** in 1957.

生物学 (Biology)

3. **Lower** performance on *some* benchmarks

Model	**NQ**
BM25	59.1
DPR	**79.4**

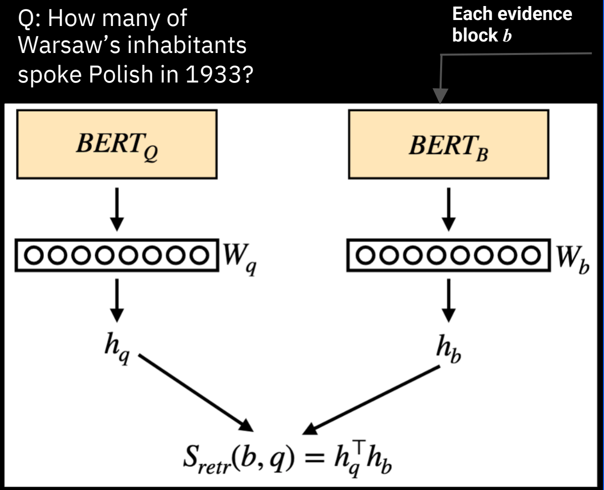
Dense Passage Retrieval (DPR)

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

Retrieve top-k passages

Perform Retrieval Methods:

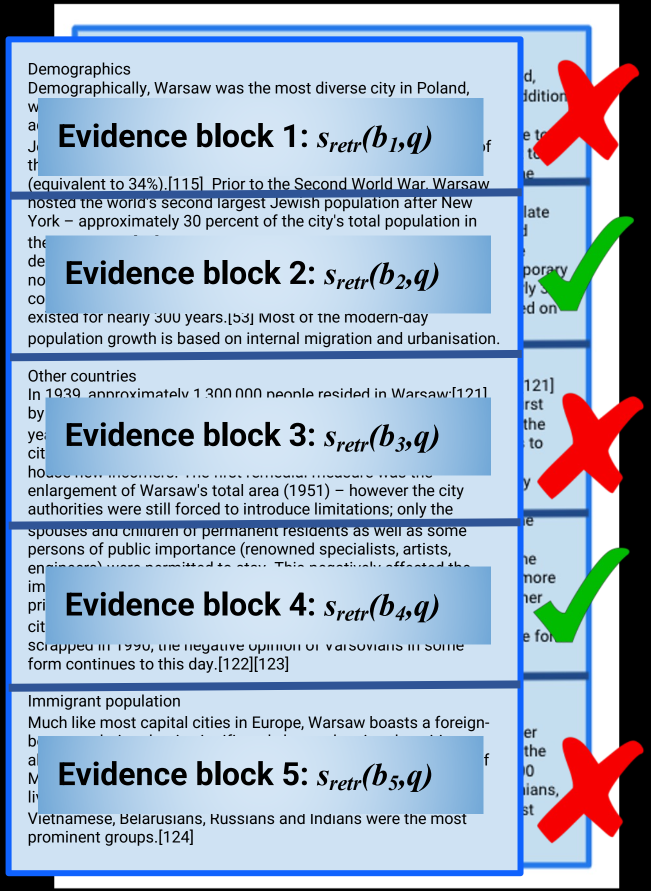
- **Neural: Dense Passage Retrieval (DPR) [Karpukhin et al., 2020]**



Retriever score:

$$
\begin{aligned}
h_q &= \mathbf{W_q}\text{BERT}_Q(q)[\texttt{CLS}] \\
h_b &= \mathbf{W_b}\text{BERT}_B(b)[\texttt{CLS}] \\
S_{retr}(b,q) &= h_q^\top h_b
\end{aligned}
$$

Select top-k blocks from collection (e.g. Wikipedia)



ColBERT

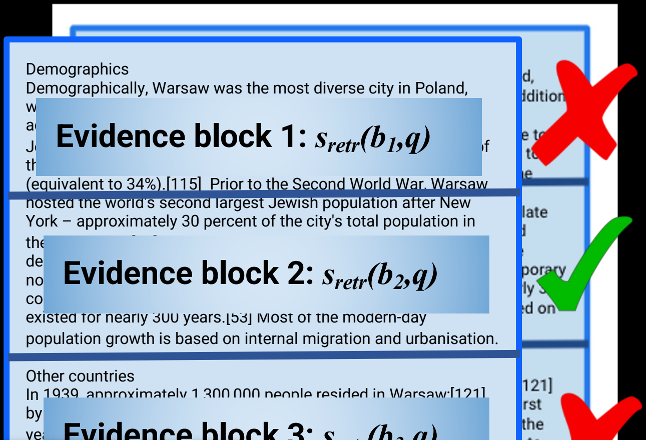
Q: How many of Warsaw's inhabitants spoke Polish in 1933?

Retrieve top-k passages

Perform Retrieval Methods:

- **Neural: ColBERT** [Khattab et al., 2021]

Select top-k blocks from collection (e.g. Wikipedia)



Demographics

Demographically, Warsaw was the most diverse city in Poland,

Evidence block 1: $s_{retr}(b_1,q)$

(equivalent to 34%).[115] Prior to the Second World War, Warsaw hosted the world's second largest Jewish population after New York – approximately 30 percent of the city's total population in

Evidence block 2: $s_{retr}(b_2,q)$

existed for nearly 300 years.[53] Most of the modern-day population growth is based on internal migration and urbanisation.

Other countries

In 1939, approximately 1,300,000 people resided in Warsaw;[121]

Evidence block 3: $s_{retr}(b_3,q)$

area (1951) – however the city

ntroduce limitations; only the

ent residents as well as some

enowned specialists, artists,

ck 4: $s_{retr}(b_4,q)$

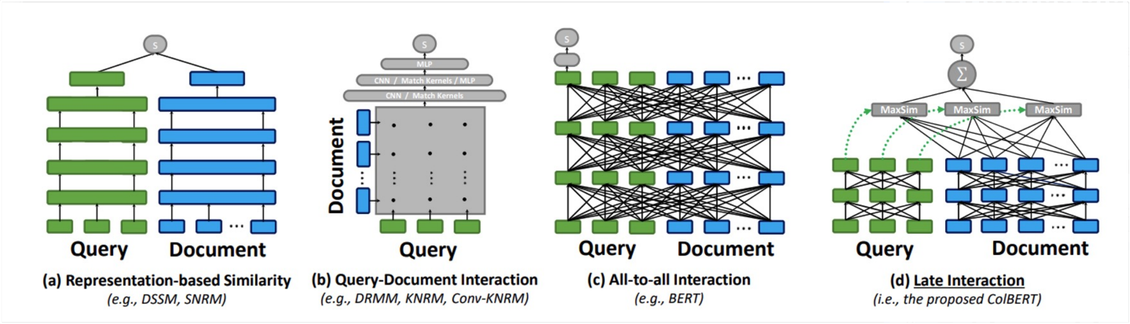
[123]

Europe, Warsaw boasts a foreign-

Evidence block 5: $s_{retr}(b_5,q)$

Vietnamese, Belarusians, Russians and Indians were the most prominent groups.[124]

Q: How many of Warsaw's inhabitants spoke Polish in 1933?



(a) Representation-based Similarity (e.g., DSSM, SNRM)

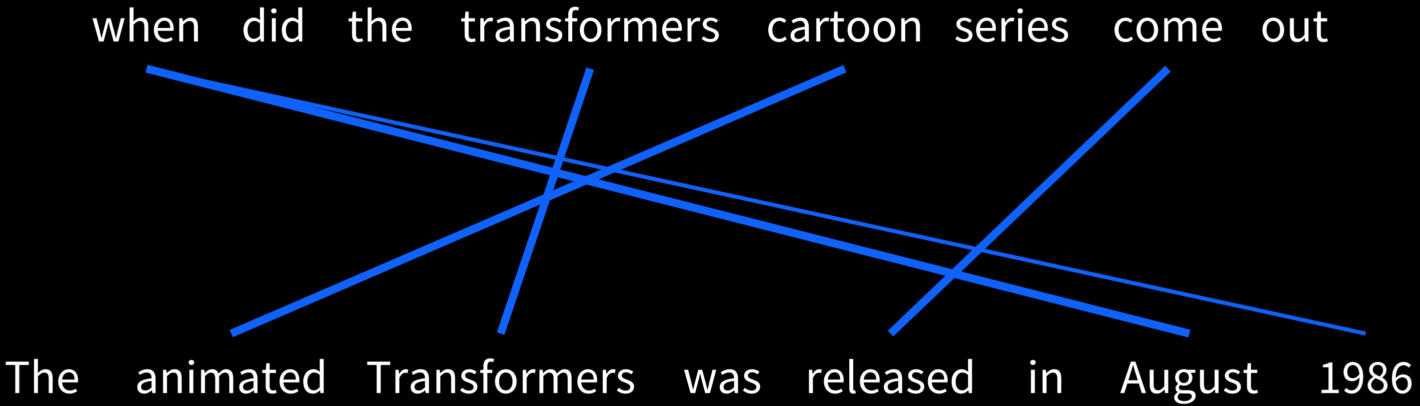
(b) Query-Document Interaction (e.g., DRMM, KNRM, Conv-KNRM)

(c) All-to-all Interaction (e.g., BERT)

(d) Late Interaction (i.e., the proposed ColBERT)

Other choices for Neural Retrievers

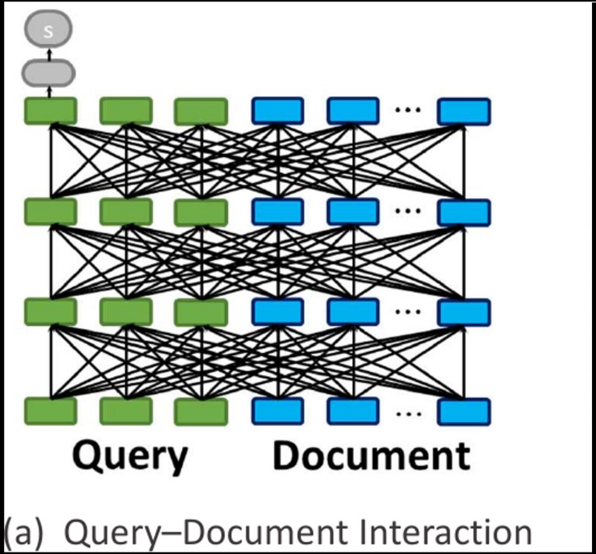
Soft-matching of query to document terms



when did the transformers cartoon series come out

The animated Transformers was released in August 1986

Situating ColBERT in the neural IR landscape

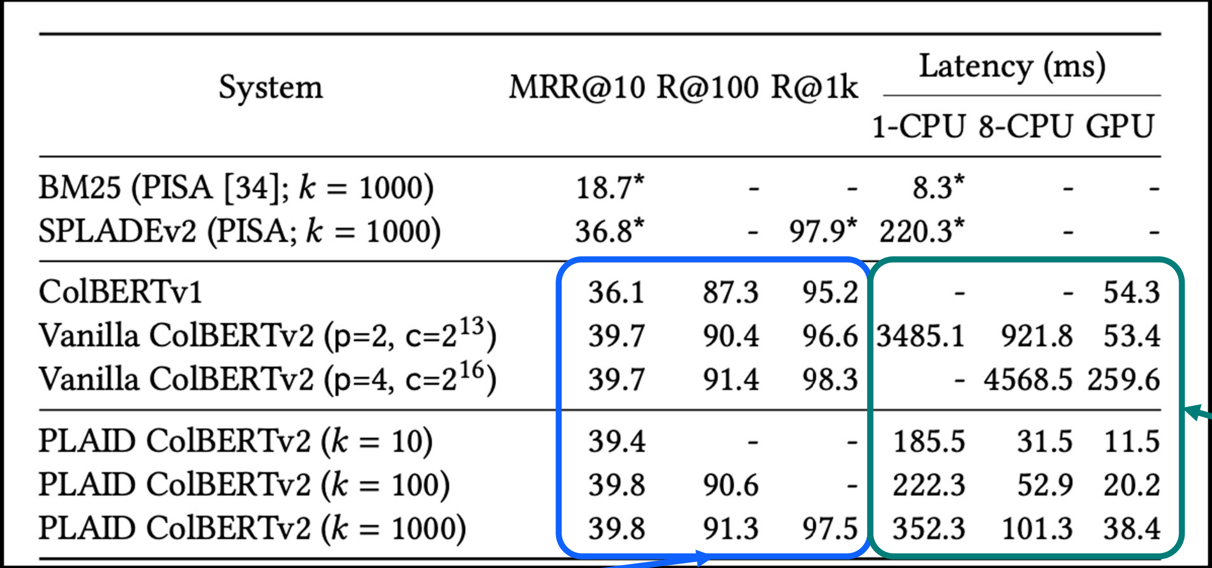


(a) Query–Document Interaction

Khattab and Zaharia 2020

PLAID ColBERT results: MS MARCO v1

PrimeQA



System	MRR@10	R@100	R@1k	Latency (ms)		
				1-CPU	8-CPU	GPU
BM25 (PISA [34]; $k = 1000$)	18.7*	-	-	8.3*	-	-
SPLADEv2 (PISA; $k = 1000$)	36.8*	-	97.9*	220.3*	-	-
ColBERTv1	36.1	87.3	95.2	-	-	54.3
Vanilla ColBERTv2 (p=2, c=2^{13})	39.7	90.4	96.6	3485.1	921.8	53.4
Vanilla ColBERTv2 (p=4, c=2^{16})	39.7	91.4	98.3	-	4568.5	259.6
PLAID ColBERTv2 ($k = 10$)	39.4	-	-	185.5	31.5	11.5
PLAID ColBERTv2 ($k = 100$)	39.8	90.6	-	222.3	52.9	20.2
PLAID ColBERTv2 ($k = 1000$)	39.8	91.3	97.5	352.3	101.3	38.4

Huge speed-ups

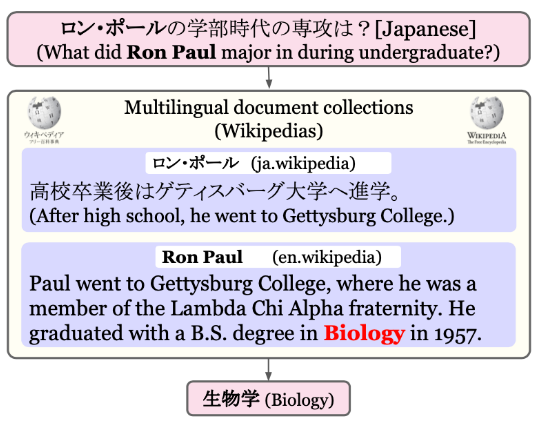
No performance loss over ColBERTv2

But do these work in other languages than English?

Problem statement

Cross-lingual Open-Retrieval Question Answering (XOR QA)

- Practical issue in QA: information scarcity and information asymmetry
- XOR QA: Enable questions from one language (non-Eng) to be answered via content from another language (English)



ロン・ポールの学部時代の専攻は？[Japanese]
(What did **Ron Paul** major in during undergraduate?)

Multilingual document collections
(Wikipedias)

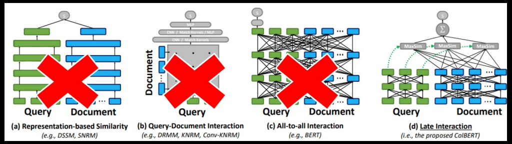
ロン・ポール (ja.wikipedia)
高校卒業後はゲティスバーグ大学へ進学。
(After high school, he went to Gettysburg College.)

Ron Paul (en.wikipedia)
Paul went to Gettysburg College, where he was a member of the Lambda Chi Alpha fraternity. He graduated with a B.S. degree in **Biology** in 1957.

生物学 (Biology)

A. Asai, 2021.naacl-main.46

Multilingual Retriever – Training Algorithm



Кака средн зарплата в Краснодаре на сегодн хны день~?

1993년 프랑스 총리는 누구인가요? (Who was the French Prime Minister in 1993?)

Krasnc … among … 0.3% c … t at addition, Krasnodar holds the first place in terms of highest average salary—21,742 rubles per capita.

Какая средняя зарплата в Краснодаре на сегодняшний день? (What is the average wage in Krasnodar?)

founded Hanei Maebashi Silk Mill and learned instrumental silk reeling techniques directly from Caspal Müller)

Tom…

governi

速水堅曹はどこで製糸技術を学んだ？ (Where did **Kenso Hayami** learn the silk-reeling technique?)

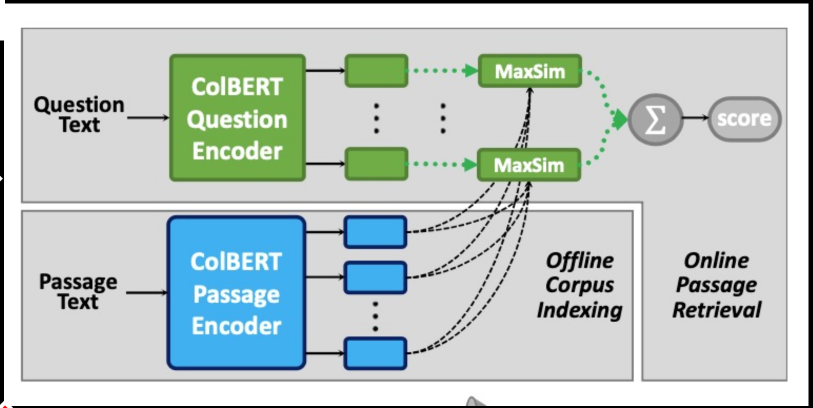
1993년 프랑스 총리는 누구인가요?

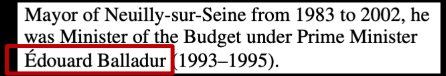
yor of Neuilly-sur-Seine from 1983 to 2002, he s Minister of the Budget under Prime Minister ouard Balladur (1993–1995).

family with five children and longstanding ties to France. His family emigrated to Marseille in the mid-to-late 1930s.

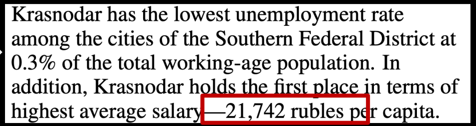
supporter of

election, it could govern with Chirac as prime minister

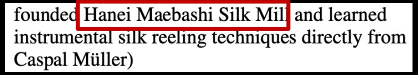




Mayor of Neuilly-sur-Seine from 1983 to 2002, he was Minister of the Budget under Prime Minister Édouard Balladur (1993–1995).



Krasnodar has the lowest unemployment rate among the cities of the Southern Federal District at 0.3% of the total working-age population. In addition, Krasnodar holds the first place in terms of highest average salary—21,742 rubles per capita.



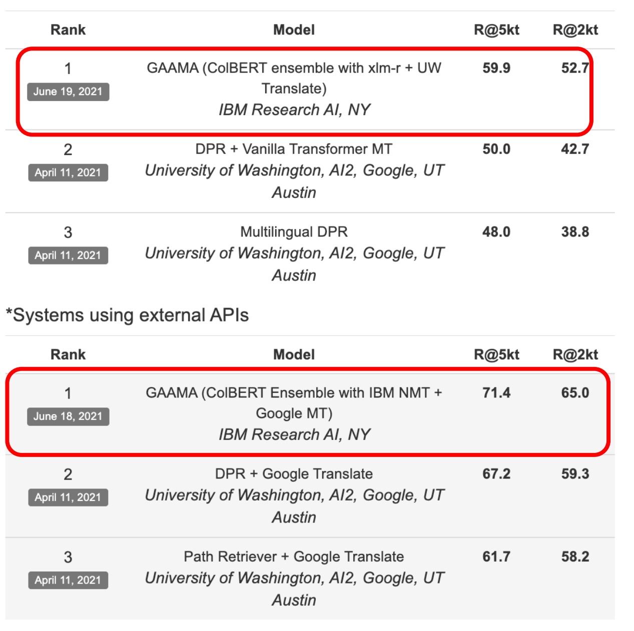
founded Hanei Maebashi Silk Mill and learned instrumental silk reeling techniques directly from Caspal Müller)

XLM-RoBERTa

BERT
(monolingual English)

IBM / Google
Translation Engine

Results



Rank	Model	R@5kt	R@2kt
1 June 19, 2021	GAAMA (ColBERT ensemble with xlm-r + UW Translate) *IBM Research AI, NY*	**59.9**	**52.7**
2 April 11, 2021	DPR + Vanilla Transformer MT *University of Washington, AI2, Google, UT Austin*	**50.0**	**42.7**
3 April 11, 2021	Multilingual DPR *University of Washington, AI2, Google, UT Austin*	**48.0**	**38.8**

*Systems using external APIs

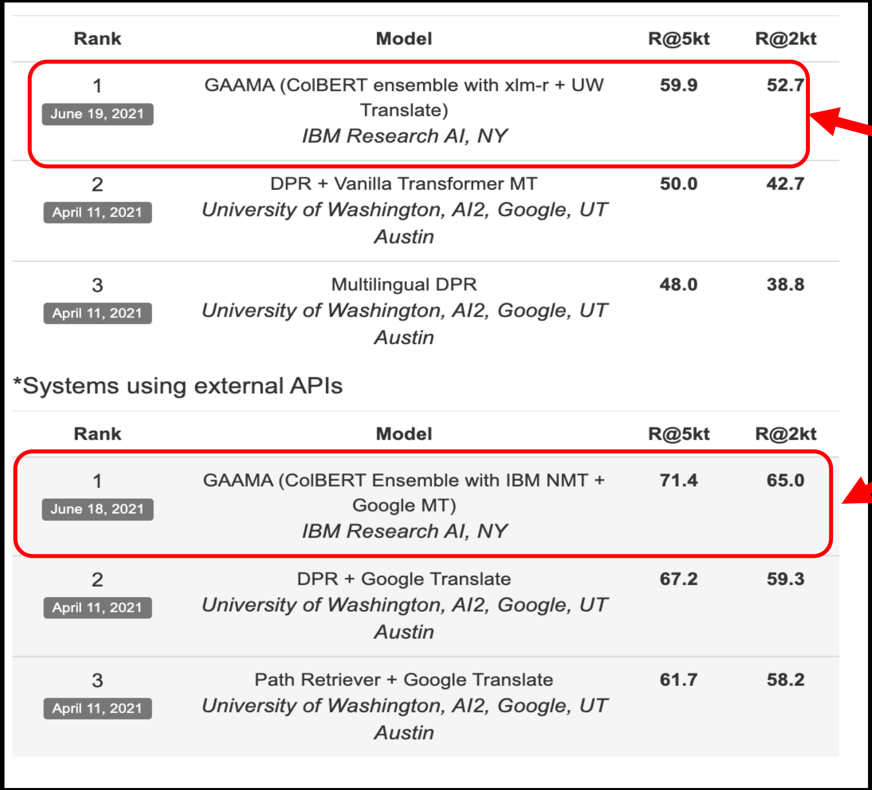
Rank	Model	R@5kt	R@2kt
1 June 18, 2021	GAAMA (ColBERT Ensemble with IBM NMT + Google MT) *IBM Research AI, NY*	**71.4**	**65.0**
2 April 11, 2021	DPR + Google Translate *University of Washington, AI2, Google, UT Austin*	**67.2**	**59.3**
3 April 11, 2021	Path Retriever + Google Translate *University of Washington, AI2, Google, UT Austin*	**61.7**	**58.2**

We obtained the top position in the leaderboard

XOR-TyDi

Cross-lingual Open-Retrieval Question Answering

Multilingual Retriever– Results on XOR TyDi Retrieve



Rank	Model	R@5kt	R@2kt
1 June 19, 2021	GAAMA (ColBERT ensemble with xlm-r + UW Translate) *IBM Research AI, NY*	59.9	52.7
2 April 11, 2021	DPR + Vanilla Transformer MT *University of Washington, AI2, Google, UT Austin*	50.0	42.7
3 April 11, 2021	Multilingual DPR *University of Washington, AI2, Google, UT Austin*	48.0	38.8

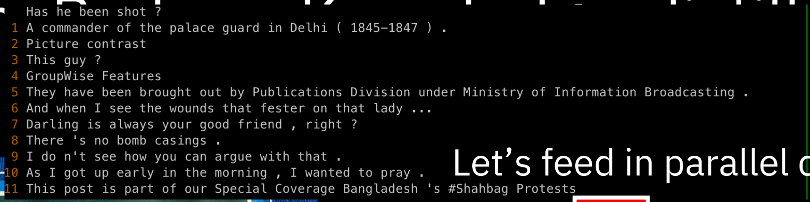
*Systems using external APIs

Rank	Model	R@5kt	R@2kt
1 June 18, 2021	GAAMA (ColBERT Ensemble with IBM NMT + Google MT) *IBM Research AI, NY*	71.4	65.0
2 April 11, 2021	DPR + Google Translate *University of Washington, AI2, Google, UT Austin*	67.2	59.3
3 April 11, 2021	Path Retriever + Google Translate *University of Washington, AI2, Google, UT Austin*	61.7	58.2

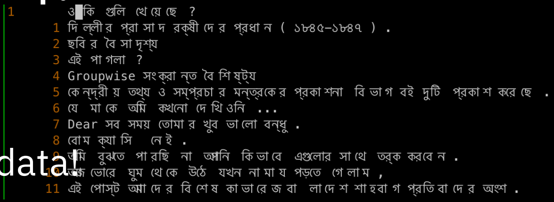
Can I make the multilingual system as good as the Monolingual system?

Then I won't need to translate my incoming queries to English!

Solution

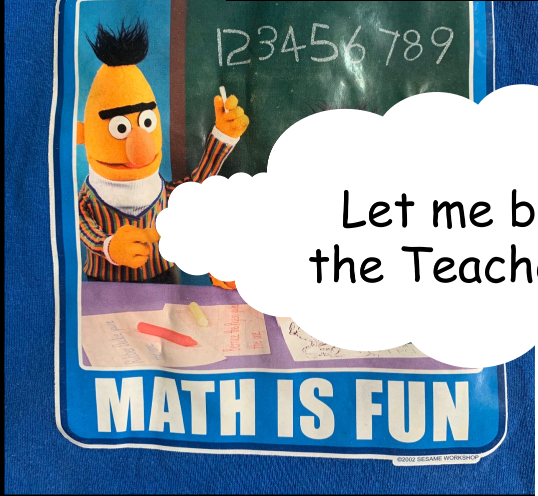


```
Has he been shot ?
A commander of the palace guard in Delhi ( 1845–1847 ) .
Picture contrast
This guy ?
GroupWise Features
They have been brought out by Publications Division under Ministry of Information Broadcasting .
And when I see the wounds that fester on that lady ...
Darling is always your good friend , right ?
There 's no bomb casings .
I do n't see how you can argue with that .
As I got up early in the morning , I wanted to pray .
This post is part of our Special Coverage Bangladesh 's #Shahbag Protests
```

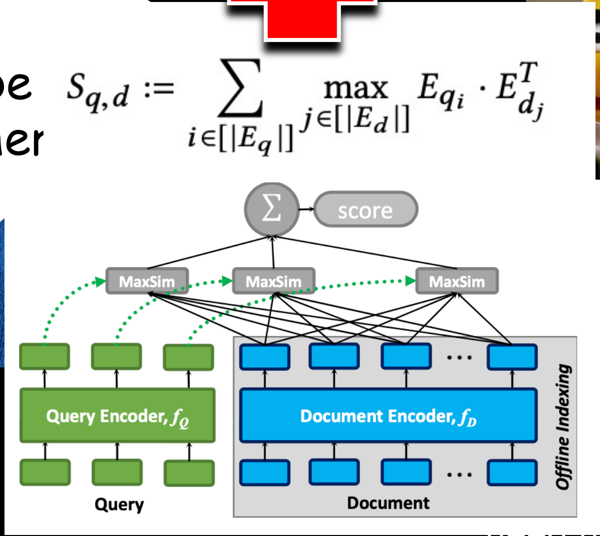


```
ও কি গুলি খেয়েছে ?
দিল্লীর প্রাসাদ রক্ষীদের প্রধান ( ১৮৪৫–১৮৪৭ ) .
ছবির বৈসাদৃশ্য
এই পাগলা ?
Groupwise সংক্রান্ত বৈশিষ্ট্য
কেন্দ্রীয় তথ্য ও সম্প্রচার মন্ত্রকের প্রকাশনা বিভাগ বই দুটি প্রকাশ করেছে .
যে মাকে আমি কখনো দেখিওনি ...
Dear সব সময় তোমার খুব ভালো বন্ধু .
বোম ক্যাসি নেই .
আমি বুঝতে পারছি না আপনি কিভাবে এগুলোর সাথে তর্ক করবেন .
আজ ভোরে ঘুম থেকে উঠে যখন নামায পড়তে গেলাম ,
এই পোস্ট আমাদের বিশেষ কাভারেজ বাংলাদেশ শাহবাগ প্রতিবাদের অংশ .
```

Let's feed in parallel data!



Let me be the Teacher



$$S_{q,d} := \sum_{i \in [|E_q|]} \max_{j \in [|E_d|]} E_{q_i} \cdot E_{d_j}^T$$

Sure! I'll be the student

Monolingual ColBERT which needs translated data

F1: 75.0 (yes, we've a better model now!)

gual ColBERT (does NOT need translated data)

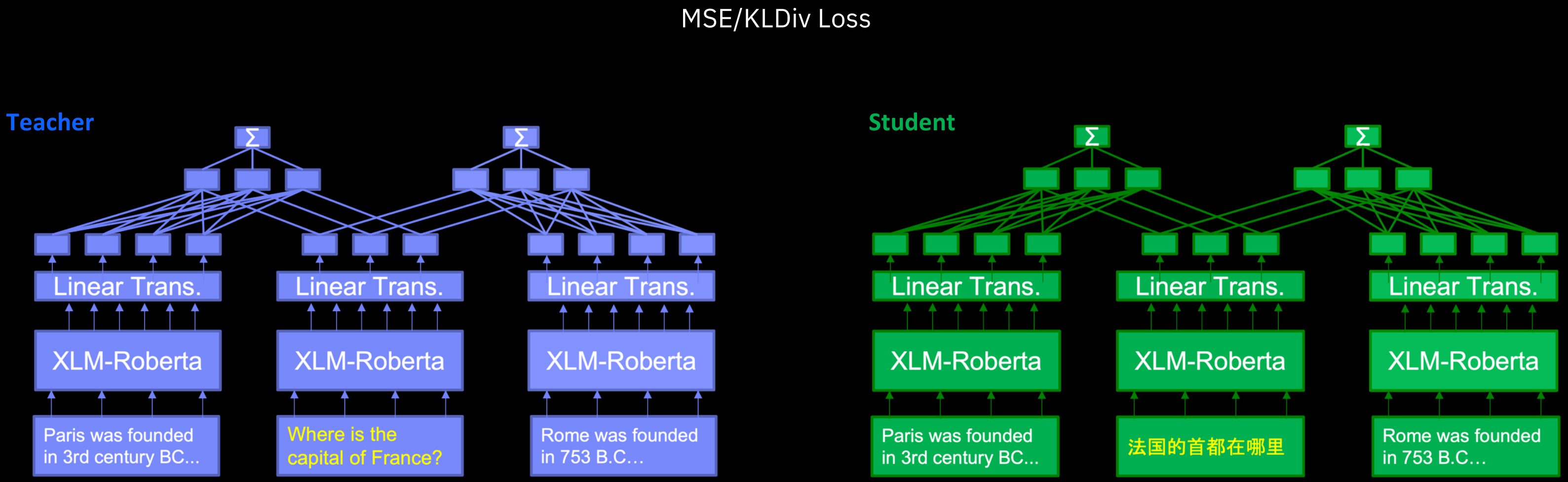
F1: 54.7

Knowledge distillation with Dr. Decr

Learning Cross-Lingual IR from an English Retriever
Li, Franz, Sultan, Iyer, Lee and Sil
NAACL, 2022

How?:

- English trained model teacher and have the cross-lingual model (student) learn from the teacher.
- Student: Dr. DECR (Dense Retrieval with Distillation-Enhanced Cross-lingual Representation)



Knowledge distillation with Dr. Decr

Idea:

– Use English trained model as teacher and have the cross-lingual model (student) learn from the teacher.

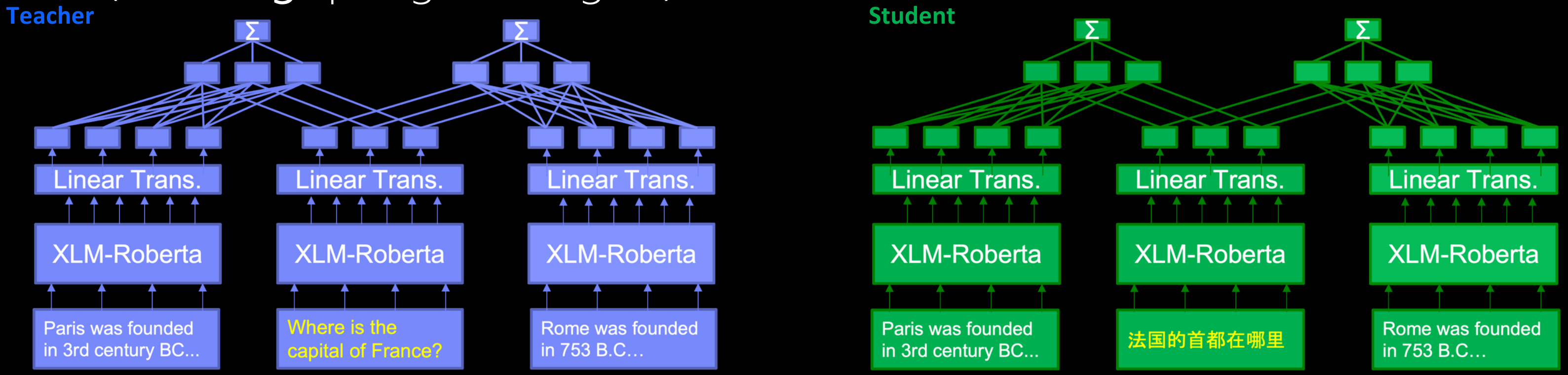
Baseline	XOR distillation	Teacher
54.9	66.1	75.1

11.2 points improvement

Input:

– Teacher: (Eng q, Eng d+, Eng d-)

– Student: (Non-Eng q, Eng d+, Eng d-)



Enhancement 1: Synthetic data

Idea:

- Using synthetic triples as extra training data in distillation

Created extra 6.5M triples

Baseline	**Baseline -> XOR distillation**	**Baseline -> Synthetic data -> XOR distillation**	**Teacher**
54.9	66.1	67.7	75.1

Another 1.6 points improvement with synthetic data

Enhancement 2: Parallel Corpus

General issue with IR dataset:

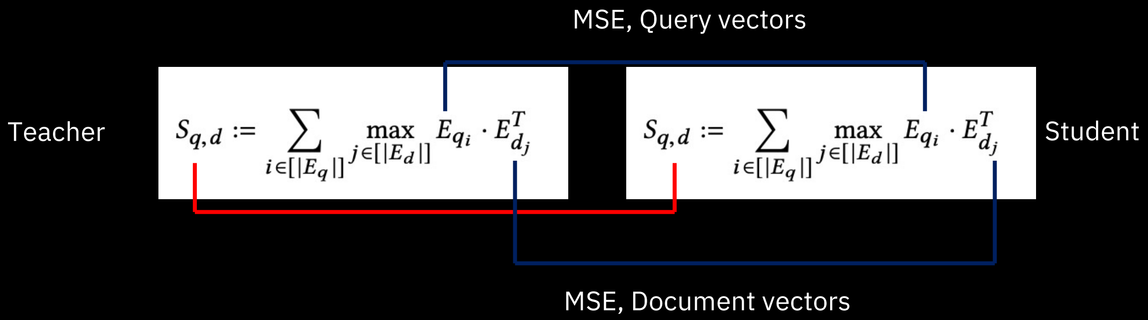
- High quality triples are difficult to make, which limits the training

Can other source of data be used to improve training?

- Parallel corpus? We have a lot!

Idea:

- Instead of teaching student to learn from teacher's score, have the student to learn from teacher's vector representation



MSE, Query vectors

Teacher

$$S_{q,d} := \sum_{i\in[|E_q|]} \max_{j\in[|E_d|]} E_{q_i} \cdot E_{d_j}^T$$

$$S_{q,d} := \sum_{i\in[|E_q|]} \max_{j\in[|E_d|]} E_{q_i} \cdot E_{d_j}^T$$

Student

MSE, Document vectors

If the student can produce same vectors as the teacher, their scores S(q,d) will also be the same

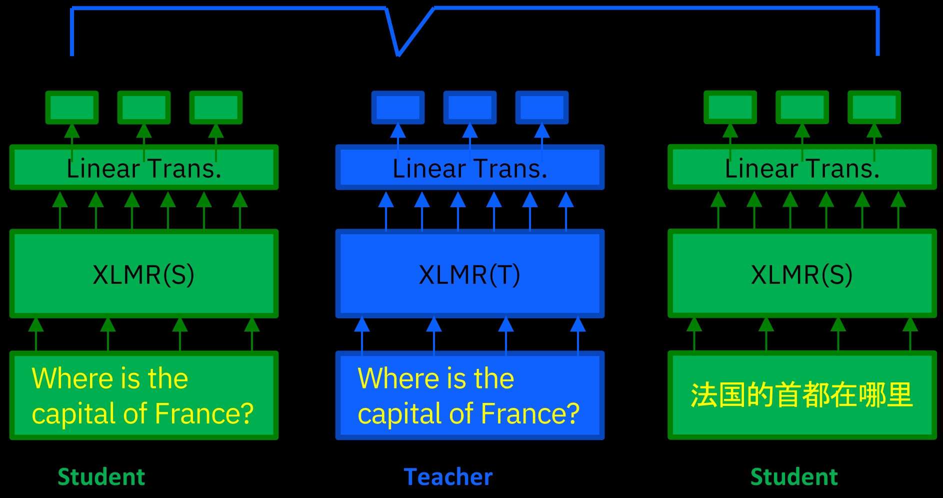
Token alignment idea

When teacher and student see different languages, which vector to learn from which?

Idea:

– Align teacher's output token with student, based on their cosine distances

– Will be noisy but hopefully can still work



Linear Trans.

XLMR(S)

Where is the capital of France?

Student

Linear Trans.

XLMR(T)

Where is the capital of France?

Teacher

Linear Trans.

XLMR(S)

法国的首都在哪里

Student

	<s>	法国	的	首都	在哪里
where	0.017	0.022	0.023	0.025	**0.014**
is	**0.014**	0.023	0.017	0.026	0.015
the	0.019	0.031	0.011	0.039	0.028
capital	0.031	0.018	0.030	**0.015**	0.019
of	0.025	0.034	**0.014**	0.043	0.032
France	0.017	**0.009**	0.026	0.021	0.016

During distillation, student sees both the Eng and Non-Eng version of the content

Summary of result

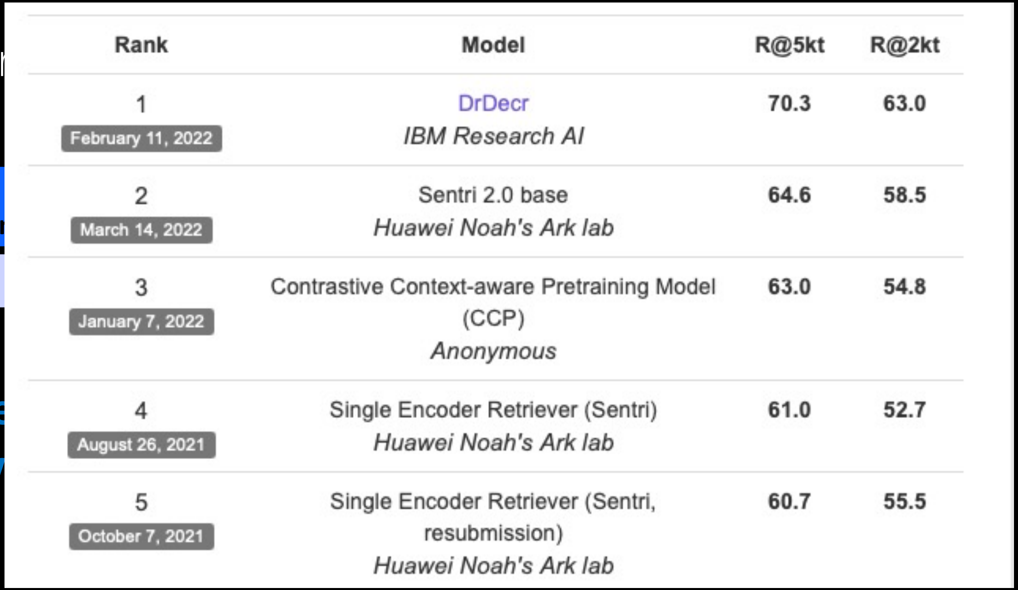
Distillation result Summary:

- 11.2 points improvement from XOR data
- 1.6 points improvement from synthetic data
- 4.4 points improvement from parallel cor

Baseline	Baseline -> XOR distillation	Baseline -> Synthetic distillatio
54.9	66.1	67.7

- **Another 4.4 points improve**
- **In total, 17.2 points improv**

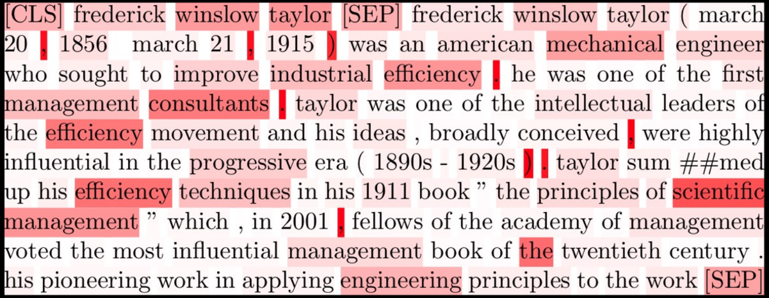
PrimeQA



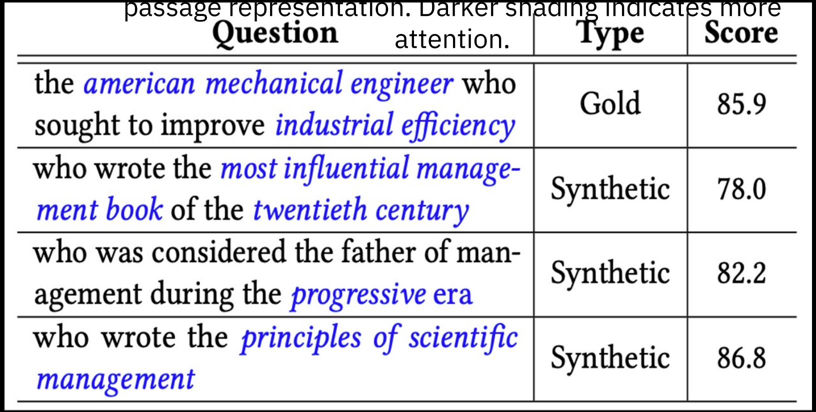
Rank	Model	R@5kt	R@2kt
1 February 11, 2022	DrDecr *IBM Research AI*	**70.3**	**63.0**
2 March 14, 2022	Sentri 2.0 base *Huawei Noah's Ark lab*	**64.6**	**58.5**
3 January 7, 2022	Contrastive Context-aware Pretraining Model (CCP) *Anonymous*	**63.0**	**54.8**
4 August 26, 2021	Single Encoder Retriever (Sentri) *Huawei Noah's Ark lab*	**61.0**	**52.7**
5 October 7, 2021	Single Encoder Retriever (Sentri, resubmission) *Huawei Noah's Ark lab*	**60.7**	**55.5**

One Limitation of Neural Retrievers

- Problem: Neural retrievers do NOT attend to many important phrases in the passage, e.g., ***academy of management*** and ***twentieth century***.
- Consequence: Low retriever scores for questions that are about these less-attended entities.
- Solution: Biases in retrievers can be overcome by generating synthetic data that is targeted towards these shortcomings.



[CLS] frederick winslow taylor [SEP] frederick winslow taylor (march 20 , 1856 march 21 , 1915) was an american mechanical engineer who sought to improve industrial efficiency . he was one of the first management consultants . taylor was one of the intellectual leaders of the efficiency movement and his ideas , broadly conceived , were highly influential in the progressive era (1890s - 1920s) . taylor sum ##med up his efficiency techniques in his 1911 book " the principles of scientific management " which , in 2001 , fellows of the academy of management voted the most influential management book of the twentieth century . his pioneering work in applying engineering principles to the work [SEP]



passage representation. Darker shading indicates more attention.

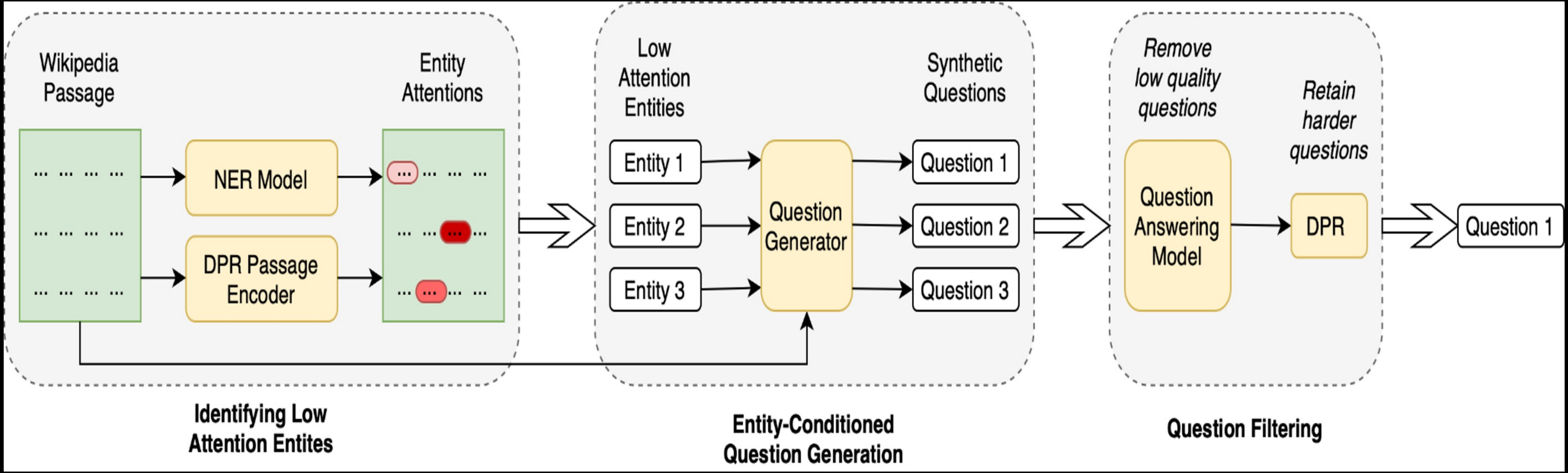
Question	Type	Score
the *american mechanical engineer* who sought to improve *industrial efficiency*	Gold	85.9
who wrote the *most influential management book* of the *twentieth century*	Synthetic	78.0
who was considered the father of management during the *progressive* era	Synthetic	82.2
who wrote the *principles of scientific management*	Synthetic	86.8

Retrieval scores from DPR for different questions corresponding to the passage in left. Important terms in the question, that are also in the passage, are shown in *italics*

Entity-Conditioned Question Generation for Robust Attention Distribution in Neural Information Retrieval.
Revanth Gangi Reddy[1], Md Arafat Sultan[2], Martin Franz[2], Avirup Sil[2], Heng Ji[1]

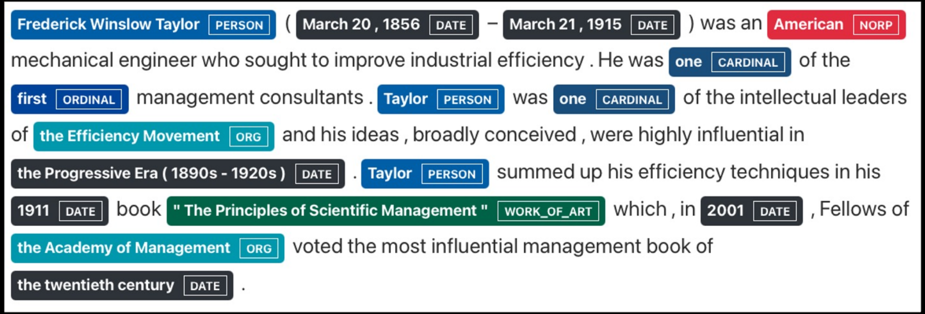
SIGIR'2022

Approach


Wikipedia Passage
NER Model
DPR Passage Encoder
Entity Attentions
Identifying Low Attention Entites
Low Attention Entities
Entity 1
Entity 2
Entity 3
Question Generator
Synthetic Questions
Question 1
Question 2
Question 3
Entity-Conditioned Question Generation
Remove low quality questions
Question Answering Model
Retain harder questions
DPR
Question 1
Question Filtering

Entity-Conditioned Question Generation

- Given a passage and an entity in that passage, we aim to generate a synthetic question about that entity.
- While training the synthetic question generator, entities within questions in existing machine reading comprehension datasets are matched against the passage to identify the conditioning entities.
- While generating synthetic IR data, entities that get lowest attentions from the IR model are used as the conditioning entities.



Frederick Winslow Taylor PERSON (March 20 , 1856 DATE – March 21 , 1915 DATE) was an American NORP mechanical engineer who sought to improve industrial efficiency . He was one CARDINAL of the first ORDINAL management consultants . Taylor PERSON was one CARDINAL of the intellectual leaders of the Efficiency Movement ORG and his ideas , broadly conceived , were highly influential in the Progressive Era (1890s - 1920s) DATE . Taylor PERSON summed up his efficiency techniques in his 1911 DATE book " The Principles of Scientific Management " WORK_OF_ART which , in 2001 DATE , Fellows of the Academy of Management ORG voted the most influential management book of the twentieth century DATE .

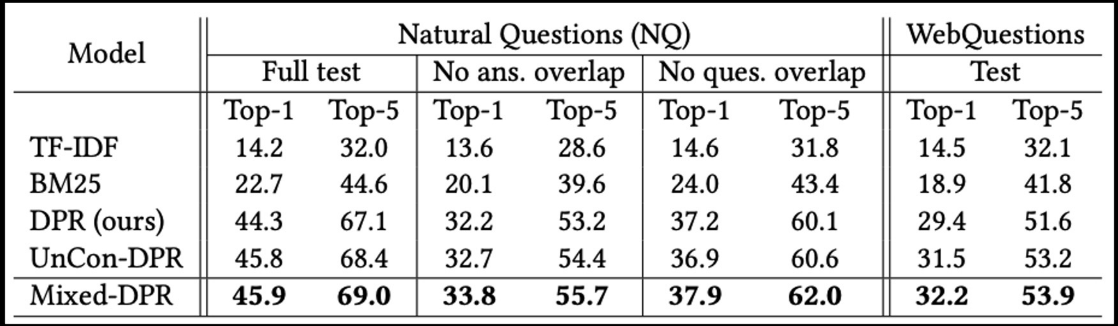
Conditioned Entity	**Generated Synthetic Question**
Progressive era	who was considered the father of management during the progressive era
Principles of Scientific Management	who wrote the principles of scientific management
Efficiency Movement	who is known as the father of efficiency movement

Experiments

- The model that uses the entity-conditioned questions within its pre-training is named *Mixed-DPR*, and is compared with the baseline DPR.

- We also compare with a model pre-trained on data that contains synthetic questions generated without any conditioning (*UnCon-DPR*).

- We see that Mixed-DPR gives upto 2% more attention to latter sentences of the passage, compared to the baseline DPR model.

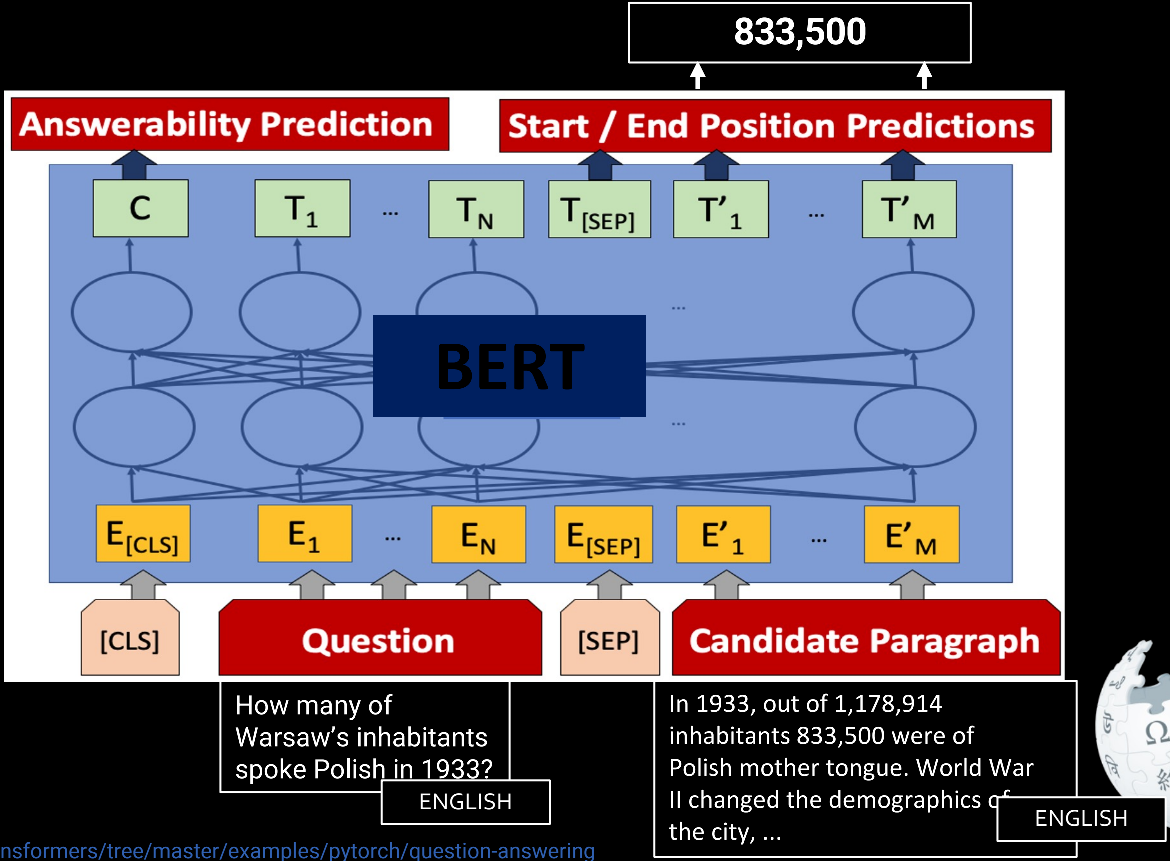


Model	Natural Questions (NQ)						WebQuestions	
	Full test		No ans. overlap		No ques. overlap		Test	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
TF-IDF	14.2	32.0	13.6	28.6	14.6	31.8	14.5	32.1
BM25	22.7	44.6	20.1	39.6	24.0	43.4	18.9	41.8
DPR (ours)	44.3	67.1	32.2	53.2	37.2	60.1	29.4	51.6
UnCon-DPR	45.8	68.4	32.7	54.4	36.9	60.6	31.5	53.2
Mixed-DPR	**45.9**	**69.0**	**33.8**	**55.7**	**37.9**	**62.0**	**32.2**	**53.9**

Reader – nuts and bolts

Machine Reading Comprehension (MRC)

- Popular choice: Add a fine-tuning layer on top of BERT [Devlin et al., 2019]



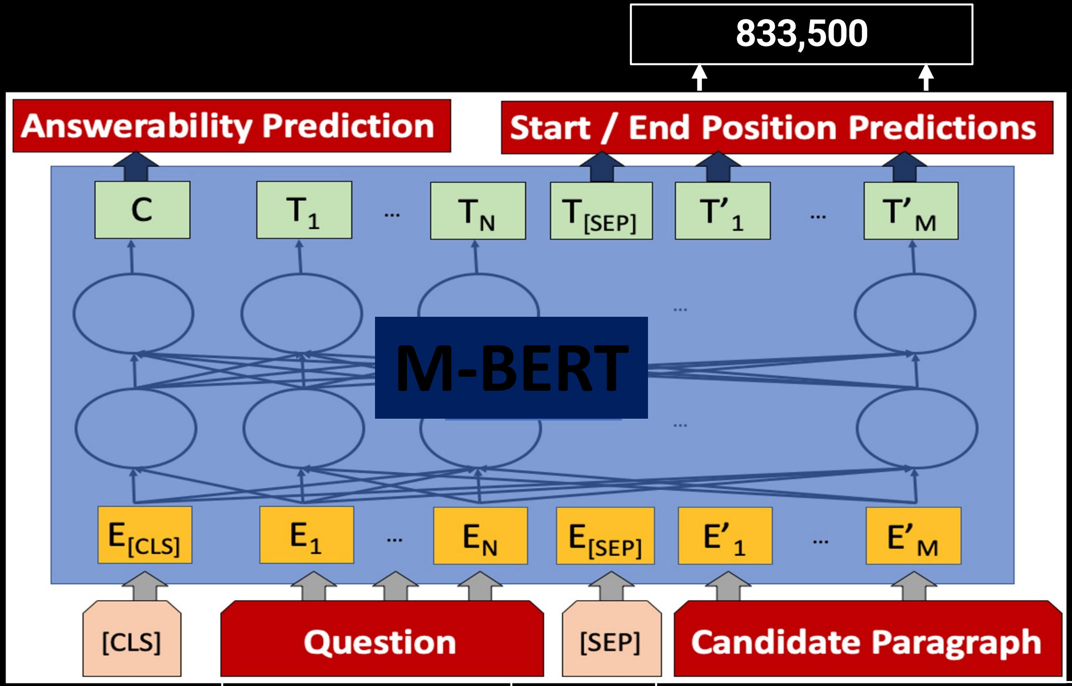
How many of Warsaw's inhabitants spoke Polish in 1933?

In 1933, out of 1,178,914 inhabitants 833,500 were of Polish mother tongue. World War II changed the demographics of the city, ...

https://github.com/huggingface/transformers/tree/master/examples/pytorch/question-answering

Multilingual Machine Reading Comprehension (MRC)

- Popular choice: Add a fine-tuning layer on top of M-BERT [Bornea et al., 2021]



1933 में वारसॉ के कितने निवासियों ने पोलिश भाषा बोली थी?

HINDI

In 1933, out of 1,178,914 inhabitants 833,500 were of Polish mother tongue. World War II changed the demographics of the city, ...

ENGLISH

https://github.com/huggingface/transformers/tree/master/examples/pytorch/question-answering

Domain Generalization: Not to Overfit or Underfit?

PrimeQA

EMNLP 2022 Paper

- *Domain Generalization in QA: Not to Overfit or Underfit the Source Domains?* Md Arafat Sultan, Avirup Sil and Radu Florian

We want our models to generalize to new, unseen domains

We have access to multiple ***source*** domains with labeled training data

But how do we train on them to do well on unseen ***target*** domains?

- Common advice: Regularize training (cross-domain); **focus is on noise**
- We say: Learn your source domains well; **focus is on signal**

How do we Learn the Source Domains Better?

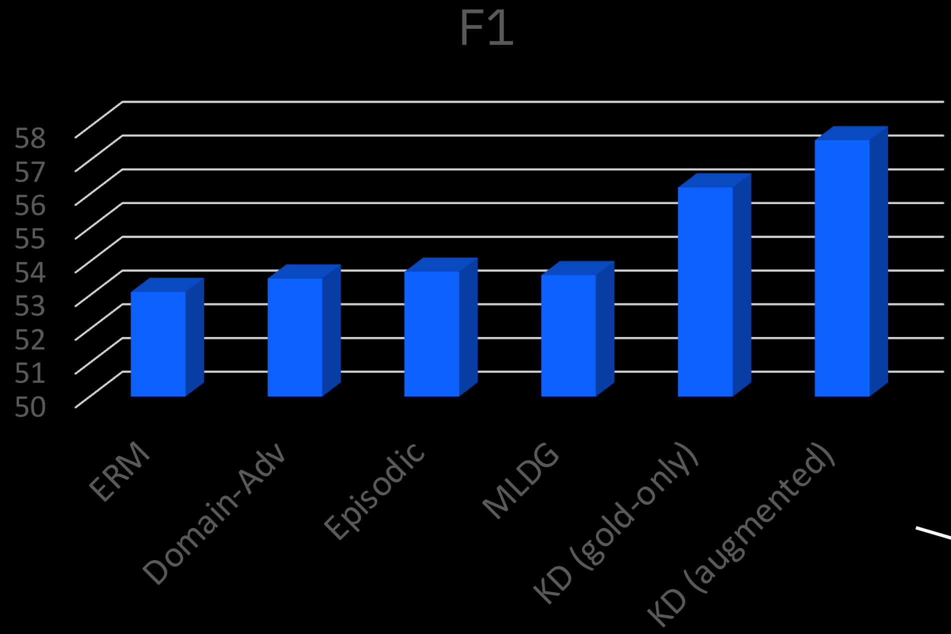
Wait, isn’t this the most basic question in machine learning? 😀

We know many ways in which to approach it

How about knowledge distillation?

- Let a bigger model learn the source domains first
  - High capacity, low inductive bias ⟹ better in-domain and OOD generalization
- Then learn from this *teacher* model, not directly from the data

Domain Generalization: Results on MRQA



MRQA (Fisch et al., 2019):

- Reading comprehension DG benchmark
- 6 source (train, dev) and 6 target (eval) datasets

"Domain-Invariant Learning" Baselines:

- Domain-Adversarial Training (Ganin et al., 2016)
- Episodic Training (Li et al., 2019)
- Meta-learning for DG (Li et al., 2018)

Not underfitting does indeed seem more important than not overfitting!

Various Setups of ORQA

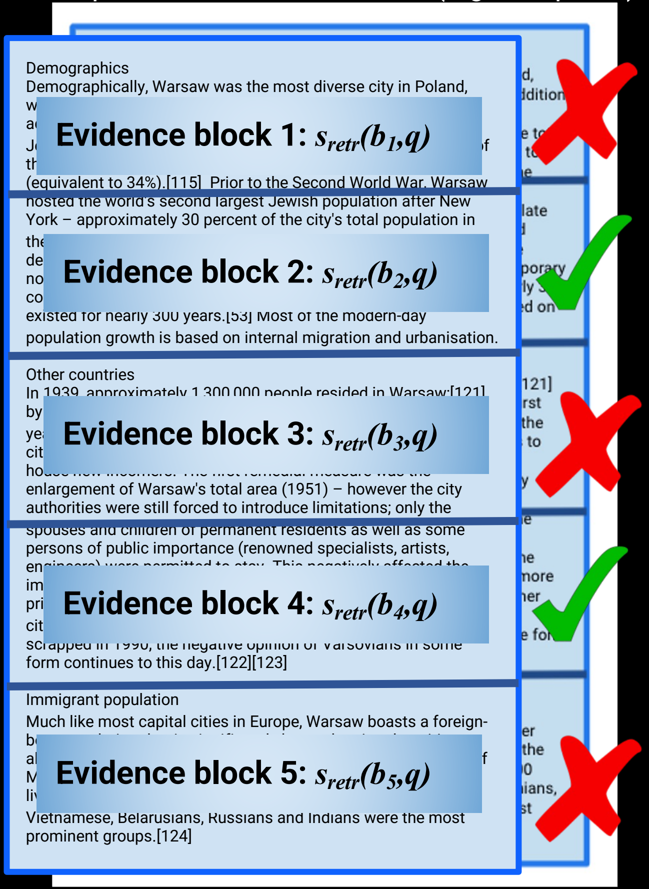
PrimeQA

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

Retriever

- DPR
- ColBERT

Select top-k blocks from collection (e.g. Wikipedia)



Demographics
Demographically, Warsaw was the most diverse city in Poland,

Evidence block 1: $s_{retr}(b_1,q)$

(equivalent to 34%).[115] Prior to the Second World War, Warsaw
hosted the world's second largest Jewish population after New
York – approximately 30 percent of the city's total population in

Evidence block 2: $s_{retr}(b_2,q)$

existed for nearly 300 years.[53] Most of the modern-day
population growth is based on internal migration and urbanisation.

Other countries
In 1939, approximately 1,300,000 people resided in Warsaw;[121]

Evidence block 3: $s_{retr}(b_3,q)$

enlargement of Warsaw's total area (1951) – however the city
authorities were still forced to introduce limitations; only the
spouses and children of permanent residents as well as some
persons of public importance (renowned specialists, artists,

Evidence block 4: $s_{retr}(b_4,q)$

scrapped in 1990, the negative opinion of Varsovians in some
form continues to this day.[122][123]

Immigrant population
Much like most capital cities in Europe, Warsaw boasts a foreign-

Evidence block 5: $s_{retr}(b_5,q)$

Vietnamese, Belarusians, Russians and Indians were the most
prominent groups.[124]

Reader

833,500

Reader

The number of inhabitants that spoke Polish in 1933 Warsaw is 833,500.

Fusion in Decoder [Izacard & Grave, 2020]

What about QA over multimedia data e.g. images & text?

MuMuQA: Multimedia Multi-hop QA

Given a news article with an image-caption pair and a question, a system needs to answer the question by extracting a short span from the body text.

Answering the questions require *multi-hop reasoning*:

- The first hop requires cross-media grounding between image and caption to get the *bridge item*.
- The second hop requires reasoning over the news body text by using the bridge item to extract the final answer.

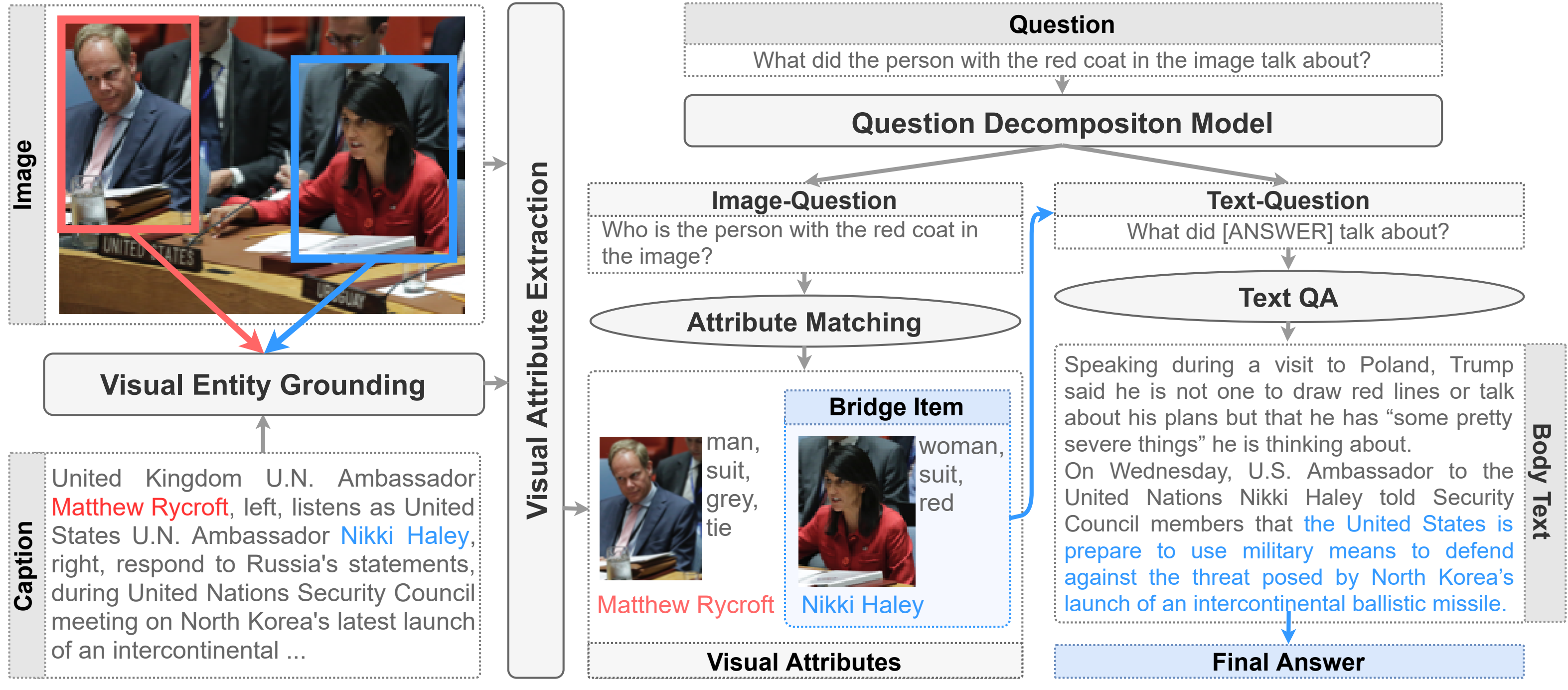
The benchmark reflects questions that news readers might have after looking at the visual information in the news article, without having read the relatively longer body text.

Image - Caption	**Body Text**
Israeli Prime Minister Benjamin Netanyahu (R) speaks with Finance Minister Moshe Kahlon during the weekly cabinet meeting in Jerusalem	A dispute between Israeli Prime Minister Benjamin Netanyahu and his finance minister over broadcast regulation sparked speculation on Sunday that Netanyahu could seek an election two years ahead of schedule. … The Israeli media quoted Netanyahu as telling ministers from his **Likud** party that he would dissolve the government if Kahlon didn't fall into line. Kahlon heads the Kulanu party, a center-right partner in Netanyahu's …

Question: What party does the person with the blue tie in the image belong to?
Answer: Likud

MuMuQA: Multimedia Multi-Hop News Question Answering via Cross-Media Knowledge Extraction and Grounding
Reddy et al., AAAI 2022

Pipeline-based Multimedia QA



Image

Visual Entity Grounding

Caption

United Kingdom U.N. Ambassador Matthew Rycroft, left, listens as United States U.N. Ambassador Nikki Haley, right, respond to Russia's statements, during United Nations Security Council meeting on North Korea's latest launch of an intercontinental ...

Visual Attribute Extraction

Question

What did the person with the red coat in the image talk about?

Question Decompositon Model

Image-Question

Who is the person with the red coat in the image?

Attribute Matching

man, suit, grey, tie

Matthew Rycroft

Bridge Item

woman, suit, red

Nikki Haley

Visual Attributes

Text-Question

What did [ANSWER] talk about?

Text QA

Speaking during a visit to Poland, Trump said he is not one to draw red lines or talk about his plans but that he has “some pretty severe things” he is thinking about.

On Wednesday, U.S. Ambassador to the United Nations Nikki Haley told Security Council members that the United States is prepare to use military means to defend against the threat posed by North Korea’s launch of an intercontinental ballistic missile.

Body Text

Final Answer

Results and Analysis

Model	**Dev**	**Test**
Multi-hop Text-only QA	25.6	24.8
End-to-end Multimedia QA	12.1	11.5
Pipeline-based Multimedia QA	**37.3**	**32.6**
Human Baseline	-	*66.5*

F1 Performance (%) of different baselines on the MuMuQA evaluation benchmark.

Caption: A woman places flowers on an altar set up in honor of Berta Caceres during a demonstration outside Honduras' embassy in Mexico City, June 15, 2016.

Question: Where was the person in the photo in the image from?

Bridge Item: Berta Caceres

An example where the grounding system failed to capture the gold bridge item (in green). The grounded entity is in blue in the caption and its corresponding bounding box is shown in blue in the image.

Coming soon to

PrimeQA

However, how can we perform these experiments easily?

- Is there a SINGLE repository that contains the latest & greatest in QA research already?
- Is that based on HuggingFace’s transformers library?

Welcome Cecilia ! A new graduate student

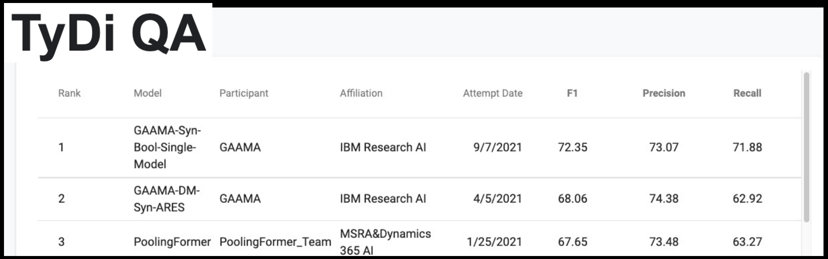
- Cecilia has taken ML 101, NLP 101 and knows basic QA details.

- She has read about: the academic benchmarks for performing QA:
  - Multilingual Machine Reading Comprehension: TyDI [Clark2019]
  - Cross-lingual Open Retrieval: XOR-TyDI [Asai2020]
  - Table QA: WikiSQL [Zhong2017]

Image source: www.phd-comics.com

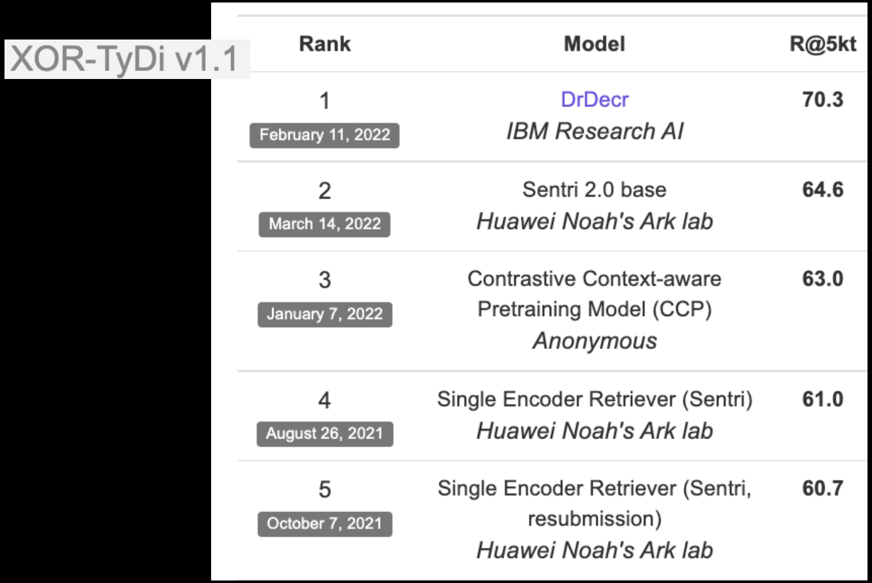
Cecilia does some literature survey!

- Cecilia wants to get the latest greatest SOTA models to start with!
  - She sees the following leaderboards ->
  - She reads the following papers: SOTA on the tasks
    - TAPAS --[Herzig2020_ACL] – SOTA on WikiSQL
    - Dr. Decr -- [Li2022_NAACL, Bornea2020_AAAI]– SOTA on XOR TyDI



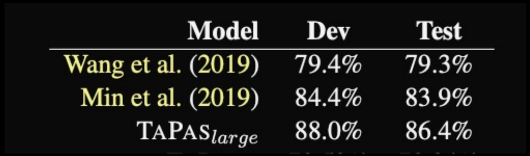
TyDi QA

Rank	Model	Participant	Affiliation	Attempt Date	F1	Precision	Recall
1	GAAMA-Syn-Bool-Single-Model	GAAMA	IBM Research AI	9/7/2021	72.35	73.07	71.88
2	GAAMA-DM-Syn-ARES	GAAMA	IBM Research AI	4/5/2021	68.06	74.38	62.92
3	PoolingFormer	PoolingFormer_Team	MSRA&Dynamics 365 AI	1/25/2021	67.65	73.48	63.27



XOR-TyDi v1.1

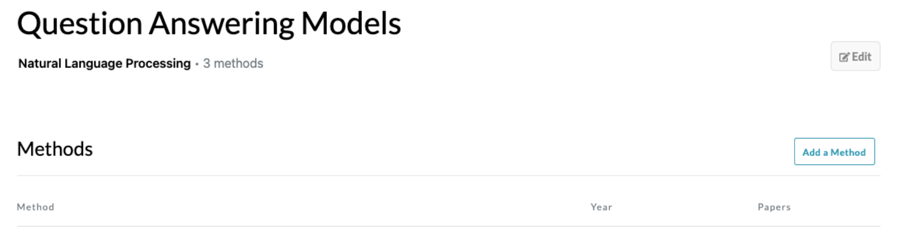
Rank	Model	R@5kt
1 February 11, 2022	DrDecr *IBM Research AI*	**70.3**
2 March 14, 2022	Sentri 2.0 base *Huawei Noah's Ark lab*	**64.6**
3 January 7, 2022	Contrastive Context-aware Pretraining Model (CCP) *Anonymous*	**63.0**
4 August 26, 2021	Single Encoder Retriever (Sentri) *Huawei Noah's Ark lab*	**61.0**
5 October 7, 2021	Single Encoder Retriever (Sentri, resubmission) *Huawei Noah's Ark lab*	**60.7**



Model	**Dev**	**Test**
Wang et al. (2019)	79.4%	79.3%
Min et al. (2019)	84.4%	83.9%
TAPAS_{large}	88.0%	86.4%

Cecilia looks for the **source code** to replicate these models

Papers with Code?



Question Answering Models

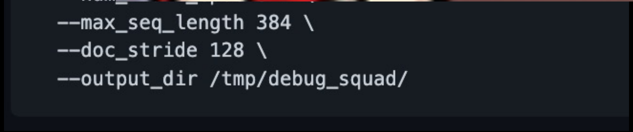
Natural Language Processing • 3 methods

Methods

Method	Year	Papers
Macaw General-Purpose Question-An		2
TransferQA Zero-Shot Dialogue State Trac		1
EMQAP Question Answering over Elec		1

👎

What about HuggingFace?



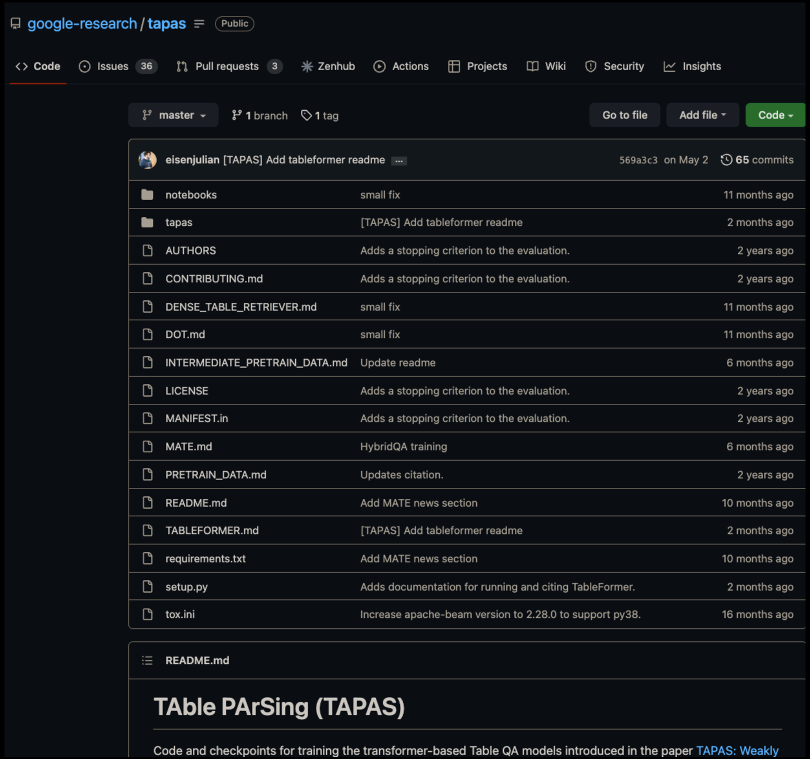
```
python
  --mod
  --dat
  --do_
  --do_
  --per
  --lea
  --num
  --max_seq_length 384 \
  --doc_stride 128 \
  --output_dir /tmp/debug_squad/
```

SQuAD2.0
The Stanford Question Answering Dataset

👎

Wait! Cecilia finds the SOTA model’s source code on TableQA

But this is only Table QA: no other QA use-case



eisenjulian [TAPAS] Add tableformer readme … 569a3c3 on May 2 65 commits

notebooks	small fix	11 months ago
tapas	[TAPAS] Add tableformer readme	2 months ago
AUTHORS	Adds a stopping criterion to the evaluation.	2 years ago
CONTRIBUTING.md	Adds a stopping criterion to the evaluation.	2 years ago
DENSE_TABLE_RETRIEVER.md	small fix	11 months ago
DOT.md	small fix	11 months ago
INTERMEDIATE_PRETRAIN_DATA.md	Update readme	6 months ago
LICENSE	Adds a stopping criterion to the evaluation.	2 years ago
MANIFEST.in	Adds a stopping criterion to the evaluation.	2 years ago
MATE.md	HybridQA training	6 months ago
PRETRAIN_DATA.md	Updates citation.	2 years ago
README.md	Add MATE news section	10 months ago
TABLEFORMER.md	[TAPAS] Add tableformer readme	2 months ago
requirements.txt	Add MATE news section	10 months ago
setup.py	Adds documentation for running and citing TableFormer.	2 months ago
tox.ini	Increase apache-beam version to 2.28.0 to support py38.	16 months ago

README.md

TAble PArSing (TAPAS)

Code and checkpoints for training the transformer-based Table QA models introduced in the paper TAPAS: Weakly

No TyDI
No XOR TyDI
No CoLBERT

Our Objective: Democratize & Replicate QA research

- We need to build a single OPEN-source repository for ALL QA problems
- End-user can use them as Lego blocks for QA problems
- End-user can modify them as per their own needs
- End-user can replicate advanced research papers and leaderboard submissions quickly

Lots of **stand-alone** Github repos

Built on top of 🤗 Hugging Face

PrimeQA

A final end2end QA solution

Bring in more repos

An alliance of QA researchers

PrimeQA has 3 basic scripts

Open-Retrieval QA (ORQA)

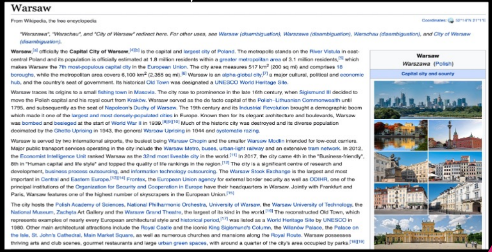
`run_qg.py`

Question Generators

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

`run_ir.py`

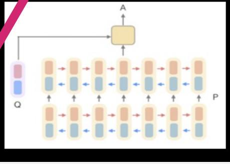
Document Retriever



`run_mrc.py`

Document Reader

833,500



General multilingual QG:

Supported Datasets:
- SQUAD
- Tydi

Table QG

Supported Datasets:
- WikiSQL

BM25

- Pyserini

Supported Datasets:
- XOR-TyDI

ColBERT

- Multilingual support (Dr.Decr)
- Knowledge Distillation

Supported Datasets:
- XOR-TyDI
- NQ

DPR

- Re-implemented to be license friendly

Supported Datasets:
- NQ
- XOR TyDI

Extractive

General MRC
(with confidence calibration):

Supported Datasets:
- TyDI
- NQ
- SQuAD v1.0
- MLQA
- XSQuAD

Special MRC:

Supported Datasets:
- Boolean QA
- List QA

With multilingual support!

These are leaderboard winners!

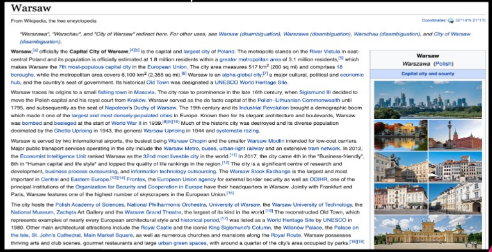
Open-Retrieval QA (ORQA)

`run_qa.py`

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

`run_ir.py`

Document Retriever



Warsaw

From Wikipedia, the free encyclopedia

Coordinates: 52°14′N 21°1′E

"Warszawa", "Warschau", and "City of Warsaw" redirect here. For other uses, see Warsaw (disambiguation), Warszawa (disambiguation), Warschau (disambiguation), and City of Warsaw (disambiguation).

Warsaw,[a] officially the **Capital City of Warsaw**,[4][b] is the capital and largest city of Poland. The metropolis stands on the River Vistula in east-central Poland and its population is officially estimated at 1.8 million residents within a greater metropolitan area of 3.1 million residents,[5] which makes Warsaw the 7th most-populous capital city in the European Union. The city area measures 517 km^2 (200 sq mi) and comprises 18 boroughs, while the metropolitan area covers 6,100 km^2 (2,355 sq mi).[6] Warsaw is an alpha-global city,[7] a major cultural, political and economic hub, and the country's seat of government. Its historical Old Town was designated a UNESCO World Heritage Site.

Warsaw traces its origins to a small fishing town in Masovia. The city rose to prominence in the late 16th century, when Sigismund III decided to move the Polish capital and his royal court from Kraków. Warsaw served as the de facto capital of the Polish–Lithuanian Commonwealth until 1795, and subsequently as the seat of Napoleon's Duchy of Warsaw. The 19th century and its Industrial Revolution brought a demographic boom which made it one of the largest and most densely-populated cities in Europe. Known then for its elegant architecture and boulevards, Warsaw was bombed and besieged at the start of World War II in 1939.[8][9][10] Much of the historic city was destroyed and its diverse population decimated by the Ghetto Uprising in 1943, the general Warsaw Uprising in 1944 and systematic razing.

Warsaw is served by two international airports, the busiest being Warsaw Chopin and the smaller Warsaw Modlin intended for low-cost carriers. Major public transport services operating in the city include the Warsaw Metro, buses, urban-light railway and an extensive tram network. In 2012, the Economist Intelligence Unit ranked Warsaw as the 32nd most liveable city in the world.[11] In 2017, the city came 4th in the "Business-friendly", 8th in "Human capital and life style" and topped the quality of life rankings in the region.[12] The city is a significant centre of research and development, business process outsourcing, and information technology outsourcing. The Warsaw Stock Exchange is the largest and most important in Central and Eastern Europe.[13][14] Frontex, the European Union agency for external border security as well as ODIHR, one of the principal institutions of the Organization for Security and Cooperation in Europe have their headquarters in Warsaw. Jointly with Frankfurt and Paris, Warsaw features one of the highest number of skyscrapers in the European Union.[15]

The city hosts the Polish Academy of Sciences, National Philharmonic Orchestra, University of Warsaw, the Warsaw University of Technology, the National Museum, Zachęta Art Gallery and the Warsaw Grand Theatre, the largest of its kind in the world.[16] The reconstructed Old Town, which represents examples of nearly every European architectural style and historical period,[17] was listed as a World Heritage Site by UNESCO in 1980. Other main architectural attractions include the Royal Castle and the iconic King Sigismund's Column, the Wilanów Palace, the Palace on the Isle, St. John's Cathedral, Main Market Square, as well as numerous churches and mansions along the Royal Route. Warsaw possesses thriving arts and club scenes, gourmet restaurants and large urban green spaces, with around a quarter of the city's area occupied by parks.[18][19]

Warsaw
Warszawa (Polish)
Capital city and county

`run_qg.py`

Question Generators

General multilingual QG:

Supported Datasets:
- SQUAD
- Tydi

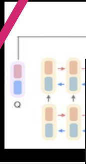
Table QG

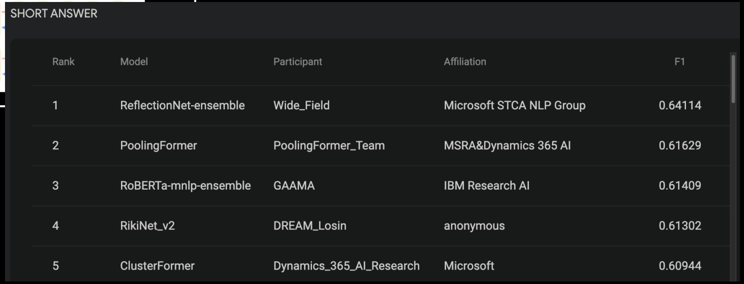
Supported Datasets:
- WikiSQL

`run_mrc.py`

Document Reader

833,500





SHORT ANSWER

Rank	Model	Participant	Affiliation	F1
1	ReflectionNet-ensemble	Wide_Field	Microsoft STCA NLP Group	0.64114
2	PoolingFormer	PoolingFormer_Team	MSRA&Dynamics 365 AI	0.61629
3	RoBERTa-mnlp-ensemble	GAAMA	IBM Research AI	0.61409
4	RikiNet_v2	DREAM_Losin	anonymous	0.61302
5	ClusterFormer	Dynamics_365_AI_Research	Microsoft	0.60944

BM25

- Pyserini

Supported
Data
- X

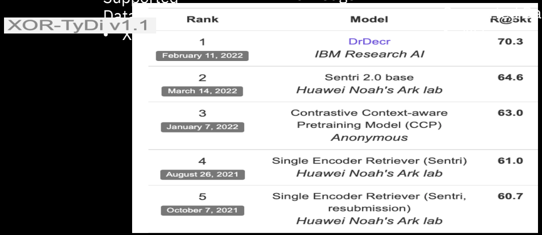
ColBERT

- Multilingual support (Dr.Decr)
- Knowledge

DPR

- Re-implemented to be license friendly

atasets:



XOR-TyDi v1.1

Rank	Model	R@5kt
1 February 11, 2022	DrDecr *IBM Research AI*	70.3
2 March 14, 2022	Sentri 2.0 base *Huawei Noah's Ark lab*	64.6
3 January 7, 2022	Contrastive Context-aware Pretraining Model (CCP) *Anonymous*	63.0
4 August 26, 2021	Single Encoder Retriever (Sentri) *Huawei Noah's Ark lab*	61.0
5 October 7, 2021	Single Encoder Retriever (Sentri, resubmission) *Huawei Noah's Ark lab*	60.7

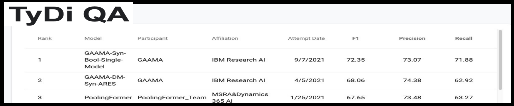
Extractive

General MRC
(with confidence calibration):

Supported Datasets:
- TyDI
- NQ
- SQuAD v1.0
- MLQA
- XSQuAD

Special MRC:

Supported Datasets:
- Boolean QA
- List QA



TyDi QA

Rank	Model	Participant	Affiliation	Attempt Date	F1	Precision	Recall
1	GAAMA-Syn-Bool-Single-Model	GAAMA	IBM Research AI	9/7/2021	72.35	73.07	71.88
2	GAAMA-DM-Syn-ARES	GAAMA	IBM Research AI	4/5/2021	68.06	74.38	62.92
3	PoolingFormer	PoolingFormer_Team	MSRA&Dynamics 365 AI	1/25/2021	67.65	73.48	63.27

PrimeQA full suite [yellow: indicates coming soon]

Open-Retrieval QA (ORQA)

`run_qg.py`

Question Generators

General multilingual QG:

Supported Datasets:
- SQUAD
- Tydi

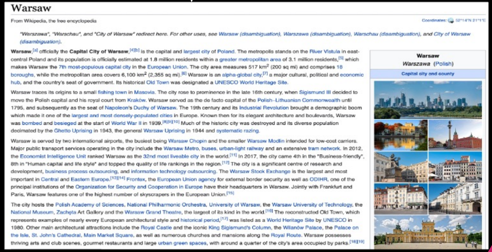
Table QG

Supported Datasets:
- WikiSQL

`run_ir.py`

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

Document Retriever



Warsaw

From Wikipedia, the free encyclopedia

Coordinates: 52°14′N 21°1′E

"Warszawa", "Warschau", and "City of Warsaw" redirect here. For other uses, see Warsaw (disambiguation), Warszawa (disambiguation), Warschau (disambiguation), and City of Warsaw (disambiguation).

Warsaw,[a] officially the **Capital City of Warsaw**,[4][b] is the capital and largest city of Poland. The metropolis stands on the River Vistula in east-central Poland and its population is officially estimated at 1.8 million residents within a greater metropolitan area of 3.1 million residents,[5] which makes Warsaw the 7th most-populous capital city in the European Union. The city area measures 517 km² (200 sq mi) and comprises 18 boroughs, while the metropolitan area covers 6,100 km² (2,355 sq mi).[6] Warsaw is an alpha-global city,[7] a major cultural, political and economic hub, and the country's seat of government. Its historical Old Town was designated a UNESCO World Heritage Site.

Warsaw traces its origins to a small fishing town in Masovia. The city rose to prominence in the late 16th century, when Sigismund III decided to move the Polish capital and his royal court from Kraków. Warsaw served as the de facto capital of the Polish–Lithuanian Commonwealth until 1795, and subsequently as the seat of Napoleon's Duchy of Warsaw. The 19th century and its Industrial Revolution brought a demographic boom which made it one of the largest and most densely-populated cities in Europe. Known then for its elegant architecture and boulevards, Warsaw was bombed and besieged at the start of World War II in 1939.[8][9][10] Much of the historic city was destroyed and its diverse population decimated by the Ghetto Uprising in 1943, the general Warsaw Uprising in 1944 and systematic razing.

Warsaw is served by two international airports, the busiest being Warsaw Chopin and the smaller Warsaw Modlin intended for low-cost carriers. Major public transport services operating in the city include the Warsaw Metro, buses, urban-light railway and an extensive tram network. In 2012, the Economist Intelligence Unit ranked Warsaw as the 32nd most liveable city in the world.[11] In 2017, the city came 4th in the "Business-friendly", 8th in "Human capital and life style" and topped the quality of life rankings in the region.[12] The city is a significant centre of research and development, business process outsourcing, and information technology outsourcing. The Warsaw Stock Exchange is the largest and most important in Central and Eastern Europe.[13][14] Frontex, the European Union agency for external border security as well as ODIHR, one of the principal institutions of the Organization for Security and Cooperation in Europe have their headquarters in Warsaw. Jointly with Frankfurt and Paris, Warsaw features one of the highest number of skyscrapers in the European Union.[15]

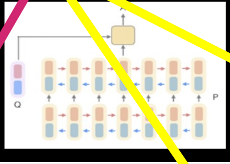
The city hosts the Polish Academy of Sciences, National Philharmonic Orchestra, University of Warsaw, the Warsaw University of Technology, the National Museum, Zachęta Art Gallery and the Warsaw Grand Theatre, the largest of its kind in the world.[16] The reconstructed Old Town, which represents examples of nearly every European architectural style and historical period,[17] was listed as a World Heritage Site by UNESCO in 1980. Other main architectural attractions include the Royal Castle and the iconic King Sigismund's Column, the Wilanów Palace, the Palace on the Isle, St. John's Cathedral, Main Market Square, as well as numerous churches and mansions along the Royal Route. Warsaw possesses thriving arts and club scenes, gourmet restaurants and large urban green spaces, with around a quarter of the city's area occupied by parks.[18][19]

Warsaw
Warszawa (Polish)
Capital city and county

`run_mrc.py`

Document Reader

833,500



BM25

- Pyserini

Supported Datasets:
- XOR-TyDI

ColBERT

- Multilingual support (Dr.Decr)
- Knowledge Distillation

Supported Datasets:
- XOR-TyDI
- NQ

DPR

- Re-implemented to be license friendly

Supported Datasets:
- NQ
- XOR TyDI

Extractive

General MRC **(with confidence calibration):**

Supported Datasets:
- TyDI
- NQ
- SQuAD v1.0
- MLQA
- XSQuAD

Special MRC:

Supported Datasets:
- Boolean QA
- List QA

Generative

General MRC:

Supported Datasets:
- NQ (KiLT)
- WoW (KiLT)

LFQA (FiD):

Supported Datasets:
- ELi5 (KilT)

Multi-modal

TableQA:

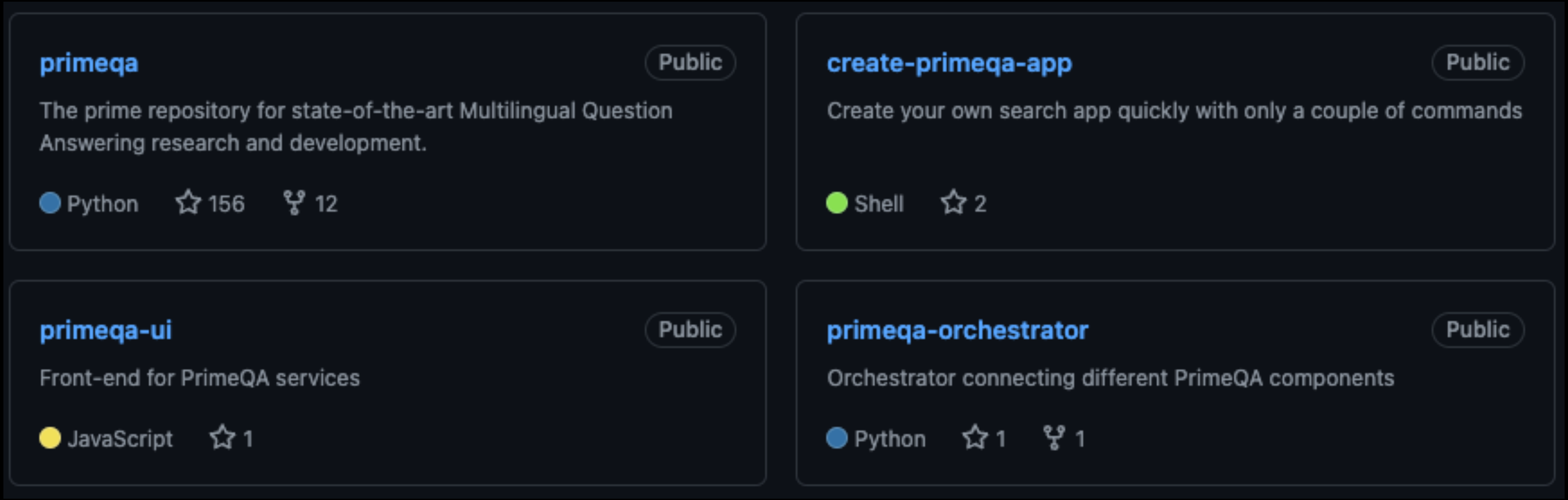
Supported Datasets:
- WikiSQL
- WikiTableQuestions

Vision + Text:

Supported Datasets:
- MumuQA

PrimeQA

- https://github.com/primeqa



primeqa Public

The prime repository for state-of-the-art Multilingual Question Answering research and development.

● Python ☆ 156 ⑂ 12

create-primeqa-app Public

Create your own search app quickly with only a couple of commands

● Shell ☆ 2

primeqa-ui Public

Front-end for PrimeQA services

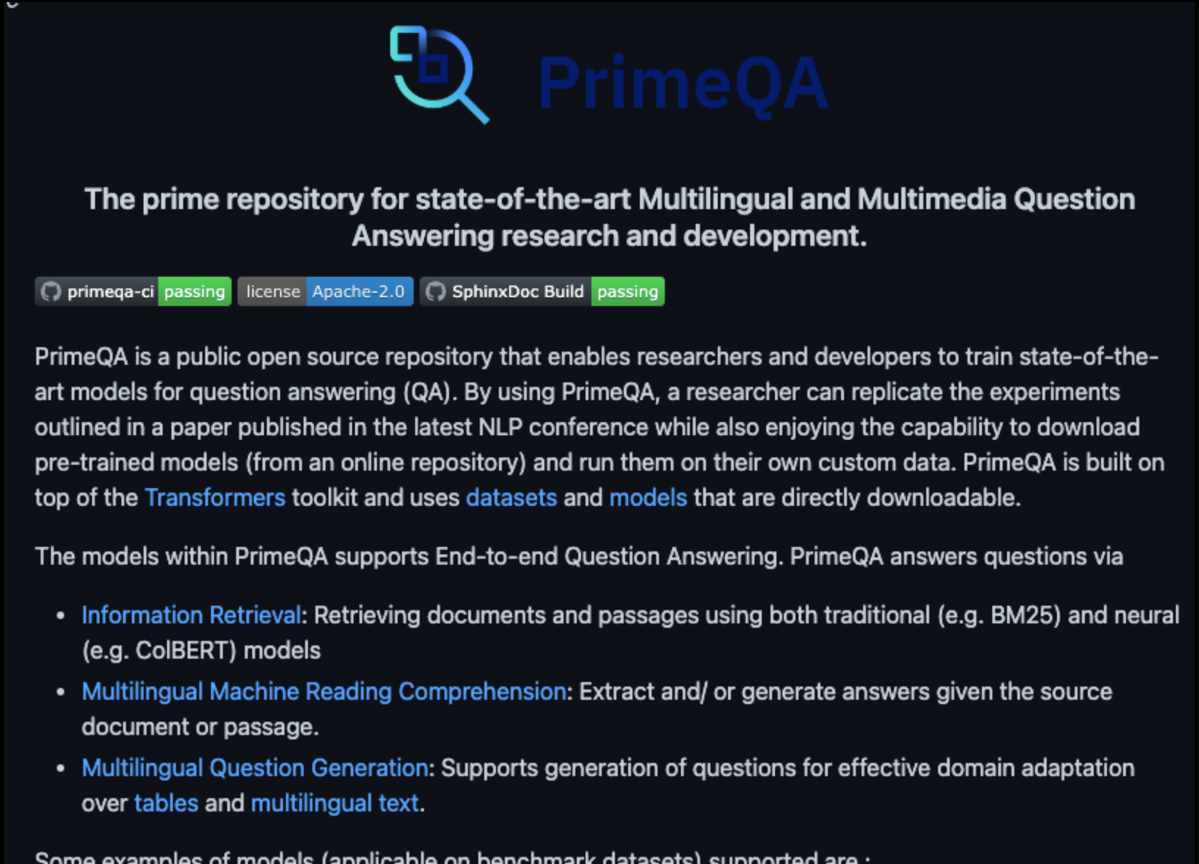
● JavaScript ☆ 1

primeqa-orchestrator Public

Orchestrator connecting different PrimeQA components

● Python ☆ 1 ⑂ 1

https://github.com/primeqa/primeqa



PrimeQA

The prime repository for state-of-the-art Multilingual and Multimedia Question Answering research and development.

primeqa-ci passing | license Apache-2.0 | SphinxDoc Build passing

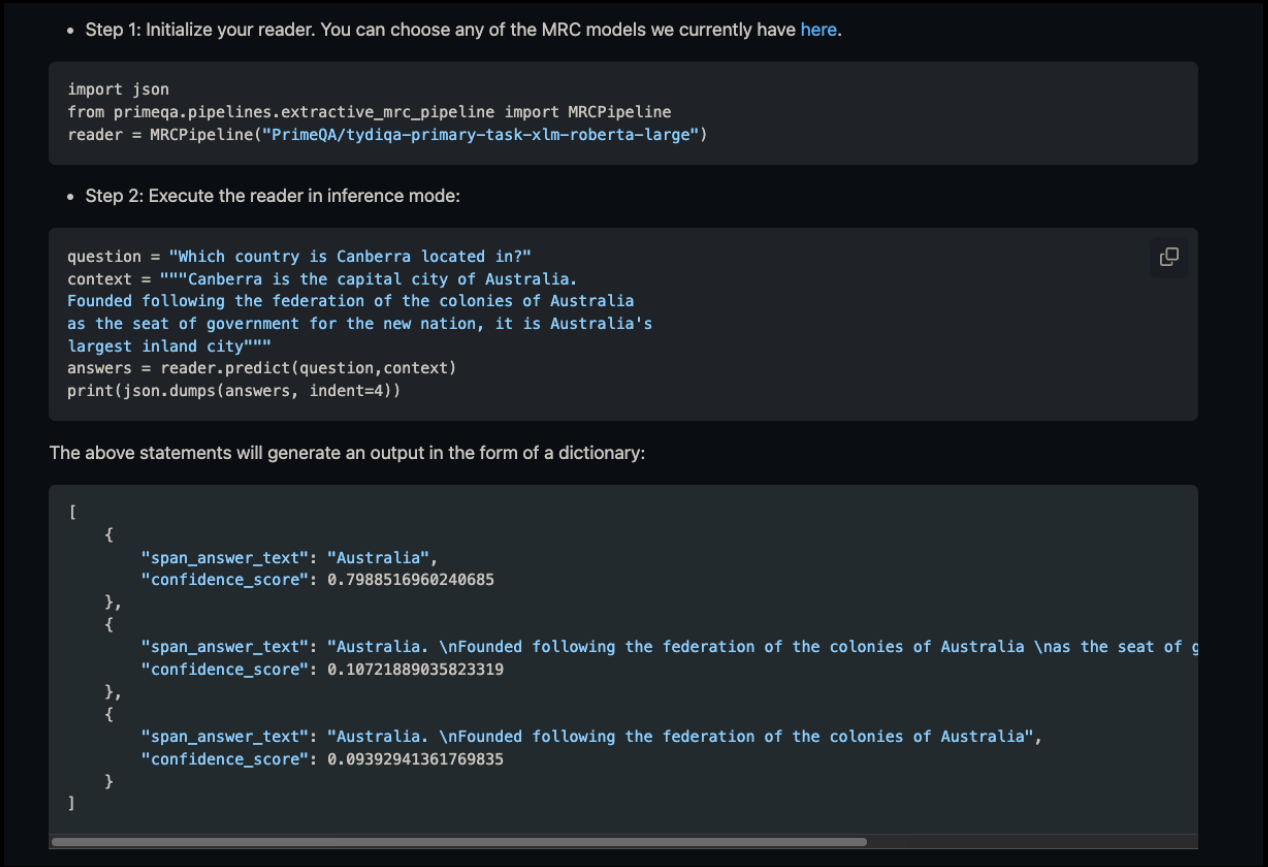
PrimeQA is a public open source repository that enables researchers and developers to train state-of-the-art models for question answering (QA). By using PrimeQA, a researcher can replicate the experiments outlined in a paper published in the latest NLP conference while also enjoying the capability to download pre-trained models (from an online repository) and run them on their own custom data. PrimeQA is built on top of the Transformers toolkit and uses datasets and models that are directly downloadable.

The models within PrimeQA supports End-to-end Question Answering. PrimeQA answers questions via

- Information Retrieval: Retrieving documents and passages using both traditional (e.g. BM25) and neural (e.g. ColBERT) models
- Multilingual Machine Reading Comprehension: Extract and/ or generate answers given the source document or passage.
- Multilingual Question Generation: Supports generation of questions for effective domain adaptation over tables and multilingual text.

Some examples of models (applicable on benchmark datasets) supported are :

Running MRC (predict mode)



- Step 1: Initialize your reader. You can choose any of the MRC models we currently have here.

```
import json
from primeqa.pipelines.extractive_mrc_pipeline import MRCPipeline
reader = MRCPipeline("PrimeQA/tydiqa-primary-task-xlm-roberta-large")
```

- Step 2: Execute the reader in inference mode:

```
question = "Which country is Canberra located in?"
context = """Canberra is the capital city of Australia.
Founded following the federation of the colonies of Australia
as the seat of government for the new nation, it is Australia's
largest inland city"""
answers = reader.predict(question,context)
print(json.dumps(answers, indent=4))
```

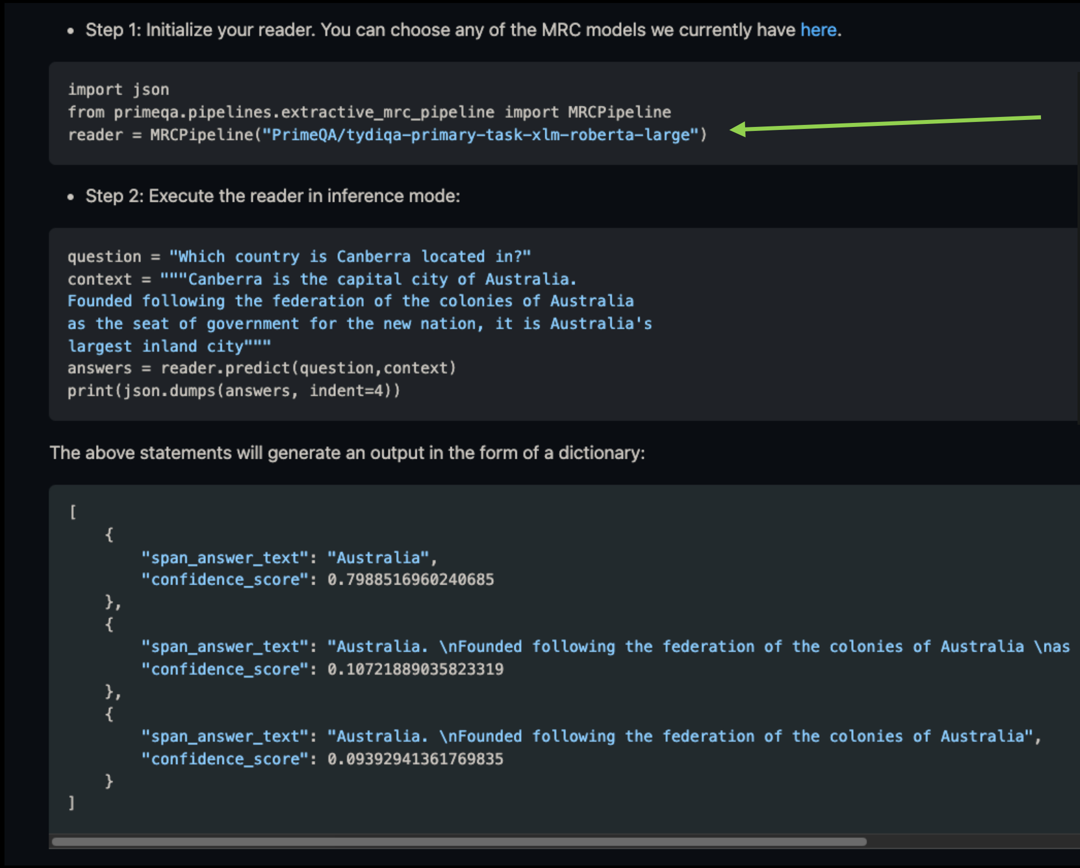
The above statements will generate an output in the form of a dictionary:

```
[
    {
        "span_answer_text": "Australia",
        "confidence_score": 0.7988516960240685
    },
    {
        "span_answer_text": "Australia. \nFounded following the federation of the colonies of Australia \nas the seat of g
        "confidence_score": 0.10721889035823319
    },
    {
        "span_answer_text": "Australia. \nFounded following the federation of the colonies of Australia",
        "confidence_score": 0.09392941361769835
    }
]
```

Inference Pipeline

- 2 PrimeQA imports,
- 4 lines of code.

Running MRC (predict mode)



- Step 1: Initialize your reader. You can choose any of the MRC models we currently have here.

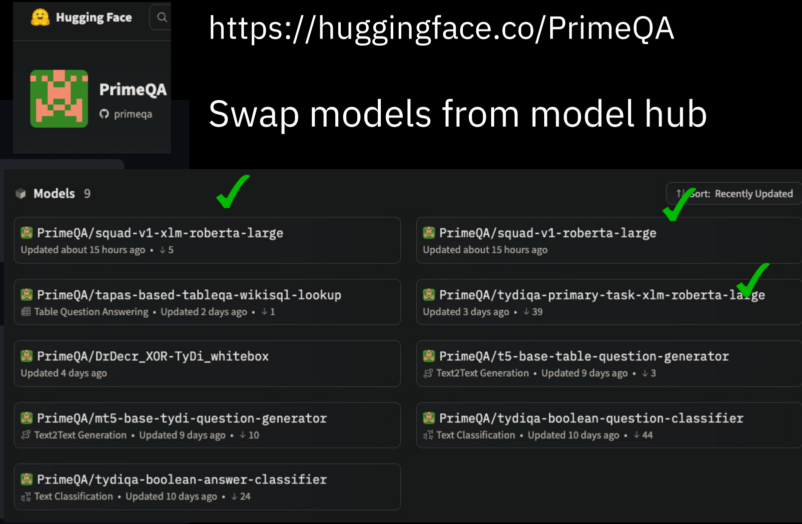
```
import json
from primeqa.pipelines.extractive_mrc_pipeline import MRCPipeline
reader = MRCPipeline("PrimeQA/tydiqa-primary-task-xlm-roberta-large")
```

- Step 2: Execute the reader in inference mode:

```
question = "Which country is Canberra located in?"
context = """Canberra is the capital city of Australia.
Founded following the federation of the colonies of Australia
as the seat of government for the new nation, it is Australia's
largest inland city"""
answers = reader.predict(question,context)
print(json.dumps(answers, indent=4))
```

The above statements will generate an output in the form of a dictionary:

```
[
    {
        "span_answer_text": "Australia",
        "confidence_score": 0.7988516960240685
    },
    {
        "span_answer_text": "Australia. \nFounded following the federation of the colonies of Australia \nas the seat of g
        "confidence_score": 0.10721889035823319
    },
    {
        "span_answer_text": "Australia. \nFounded following the federation of the colonies of Australia",
        "confidence_score": 0.09392941361769835
    }
]
```



Models 9

PrimeQA/squad-v1-xlm-roberta-large ✓
Updated about 15 hours ago • ↓5

PrimeQA/squad-v1-roberta-large ✓
Updated about 15 hours ago

PrimeQA/tapas-based-tableqa-wikisql-lookup
Table Question Answering • Updated 2 days ago • ↓1

PrimeQA/tydiqa-primary-task-xlm-roberta-large ✓
Updated 3 days ago • ↓39

PrimeQA/DrDecr_XOR-TyDi_whitebox
Updated 4 days ago

PrimeQA/t5-base-table-question-generator
Text2Text Generation • Updated 9 days ago • ↓3

PrimeQA/mt5-base-tydi-question-generator
Text2Text Generation • Updated 9 days ago • ↓10

PrimeQA/tydiqa-boolean-question-classifier
Text Classification • Updated 10 days ago • ↓44

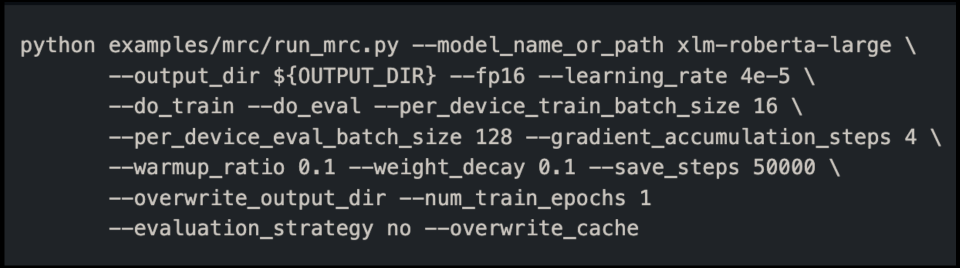
PrimeQA/tydiqa-boolean-answer-classifier
Text Classification • Updated 10 days ago • ↓24

https://huggingface.co/PrimeQA

Swap models from model hub

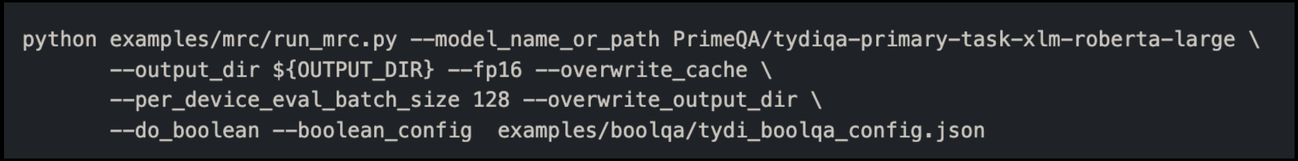
You can upload your own MRC models for others to use!

Running MRC (full train + predict)



```
python examples/mrc/run_mrc.py --model_name_or_path xlm-roberta-large \
       --output_dir ${OUTPUT_DIR} --fp16 --learning_rate 4e-5 \
       --do_train --do_eval --per_device_train_batch_size 16 \
       --per_device_eval_batch_size 128 --gradient_accumulation_steps 4 \
       --warmup_ratio 0.1 --weight_decay 0.1 --save_steps 50000 \
       --overwrite_output_dir --num_train_epochs 1
       --evaluation_strategy no --overwrite_cache
```

If your dataset has support for Boolean Questions (e.g. Yes/No) as in TyDI QA you can further run:



```
python examples/mrc/run_mrc.py --model_name_or_path PrimeQA/tydiqa-primary-task-xlm-roberta-large \
       --output_dir ${OUTPUT_DIR} --fp16 --overwrite_cache \
       --per_device_eval_batch_size 128 --overwrite_output_dir \
       --do_boolean --boolean_config  examples/boolqa/tydi_boolqa_config.json
```

Parameter to run Boolean questions

Running IR

- Training
- Indexing
- Search

```
python examples/ir/run_ir.py
  --engine_type ColBERT
  --do_index
  --compression_level 2
  --model_name_or_path <my_dir>/trained_model
  --collection <data_dir>/psgs_w100.tsv
```

```
python examples/ir/run_ir.py
  --engine_type ColBERT
  --do_train
  --triples
<data_dir>/xorqa.train_ir_negs_100_poss_3.tsv
  --model_type xlm-roberta-base
  --root <my_dir>/experiments
  --experiment <my_expt>
```

```
python examples/ir/run_ir.py
  --engine_type ColBERT
  --do_search
  --queries <data_dir>/xorqa_dev.tsv
  --index_location <my_expt>_indname
  --model_name_or_path
<my_dir>/experiments/<my_expt>/checkpoints/colbert-LAST.dnn
  --output_dir <my_dir>
```

Multilingual Question Generation: Usage

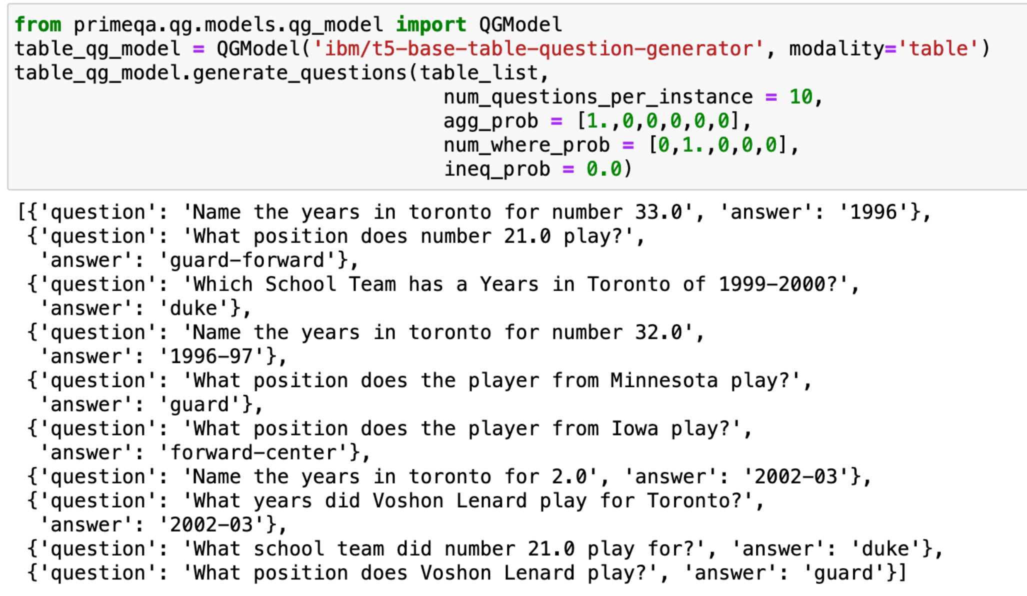
CLI – training and evaluation

```
python examples/qg/run_qg.py \
 --model_name_or_path t5-base \
 --modality passage \
 --dataset_name tydiqa \
 --do_train \
 --do_eval \
 --output_dir models/qg/$DIR_NAME \
 --learning_rate 0.0001 \
 --num_train_epochs 4\
```

CLI - generation

```
python examples/qg/run_qg.py \
 --model_name_or_path models/qg/wikisql\
 --modality table \
 --do_generate \
 --num_questions_per_instance 20 \
 --data_path <path-to-json-file> \
 --generate_aggregate \
 --max_where_clauses 2 \
 --gen_output_path /results/qg/$DIR_NAME
```

Using pretrained QG model in python code

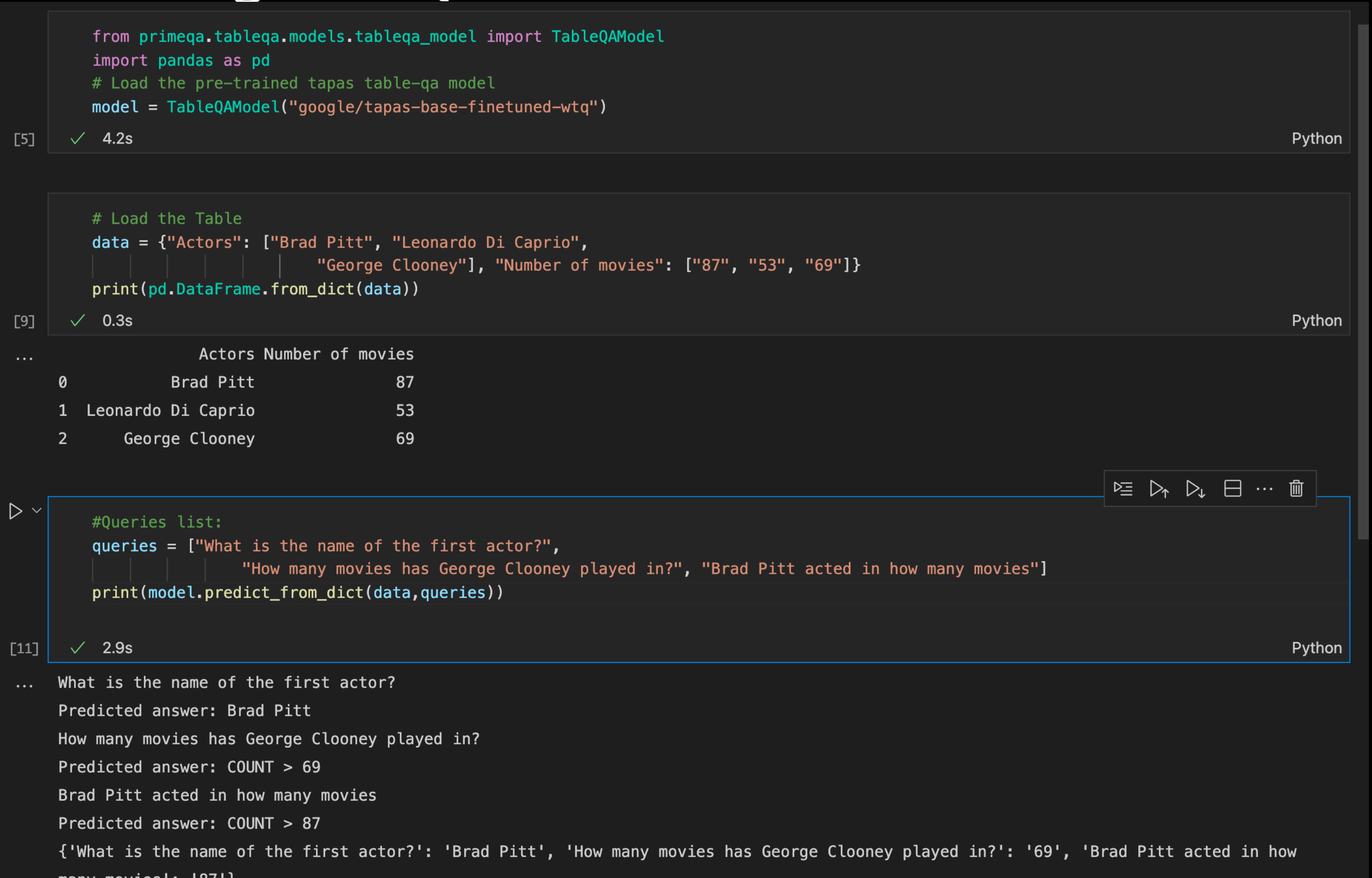


```
from primeqa.qg.models.qg_model import QGModel
table_qg_model = QGModel('ibm/t5-base-table-question-generator', modality='table')
table_qg_model.generate_questions(table_list,
                                   num_questions_per_instance = 10,
                                   agg_prob = [1.,0,0,0,0,0],
                                   num_where_prob = [0,1.,0,0,0],
                                   ineq_prob = 0.0)
```

```
[{'question': 'Name the years in toronto for number 33.0', 'answer': '1996'},
 {'question': 'What position does number 21.0 play?',
  'answer': 'guard-forward'},
 {'question': 'Which School Team has a Years in Toronto of 1999-2000?',
  'answer': 'duke'},
 {'question': 'Name the years in toronto for number 32.0',
  'answer': '1996-97'},
 {'question': 'What position does the player from Minnesota play?',
  'answer': 'guard'},
 {'question': 'What position does the player from Iowa play?',
  'answer': 'forward-center'},
 {'question': 'Name the years in toronto for 2.0', 'answer': '2002-03'},
 {'question': 'What years did Voshon Lenard play for Toronto?',
  'answer': '2002-03'},
 {'question': 'What school team did number 21.0 play for?', 'answer': 'duke'},
 {'question': 'What position does Voshon Lenard play?', 'answer': 'guard'}]
```

- ➢ One can use QG over table/passage with only 1 primeqa import and 2 code lines.

Running TableQA



```
from primeqa.tableqa.models.tableqa_model import TableQAModel
import pandas as pd
# Load the pre-trained tapas table-qa model
model = TableQAModel("google/tapas-base-finetuned-wtq")
```

```
# Load the Table
data = {"Actors": ["Brad Pitt", "Leonardo Di Caprio",
                        "George Clooney"], "Number of movies": ["87", "53", "69"]}
print(pd.DataFrame.from_dict(data))
```

```
               Actors Number of movies
0           Brad Pitt               87
1  Leonardo Di Caprio               53
2      George Clooney               69
```

```
#Queries list:
queries = ["What is the name of the first actor?",
                "How many movies has George Clooney played in?", "Brad Pitt acted in how many movies"]
print(model.predict_from_dict(data,queries))
```

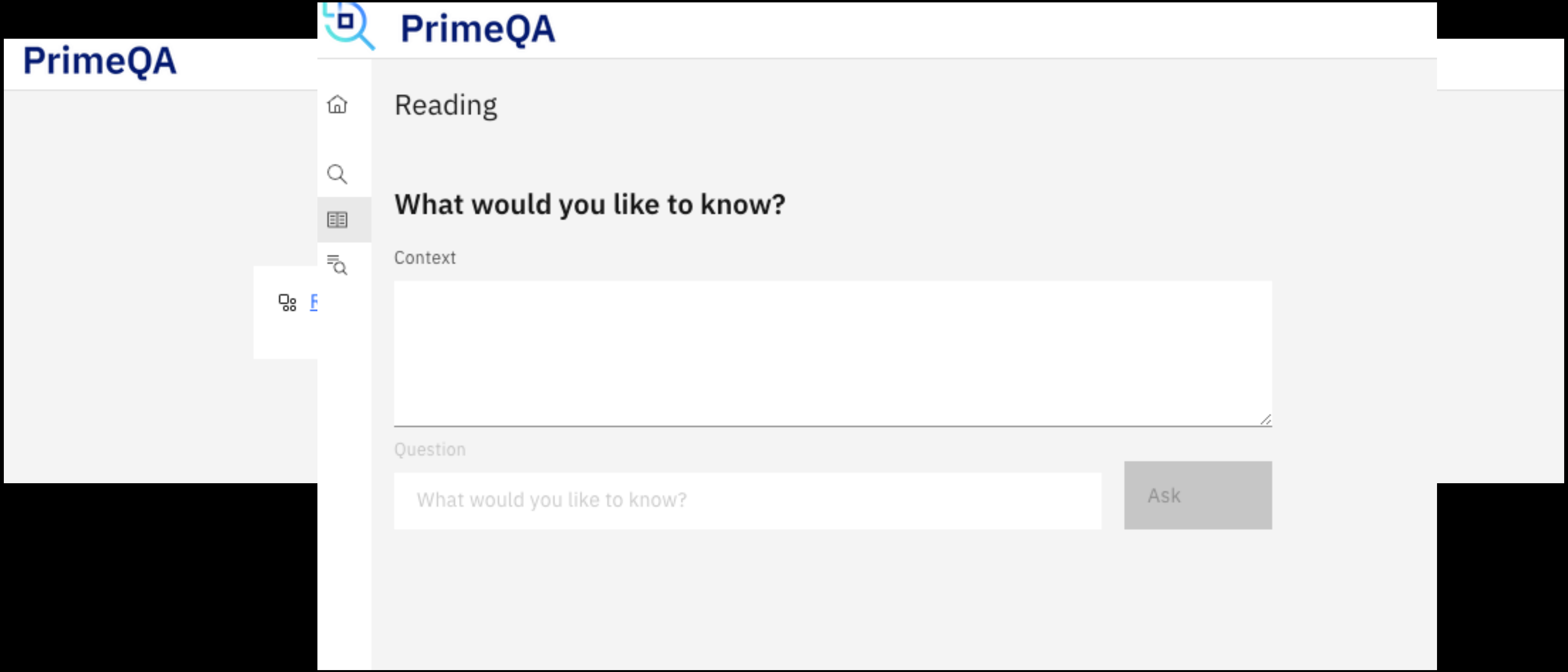
```
What is the name of the first actor?
Predicted answer: Brad Pitt
How many movies has George Clooney played in?
Predicted answer: COUNT > 69
Brad Pitt acted in how many movies
Predicted answer: COUNT > 87
{'What is the name of the first actor?': 'Brad Pitt', 'How many movies has George Clooney played in?': '69', 'Brad Pitt acted in how
many movies': '87'}
```

Inference Pipeline

- ➢ 2 PrimeQA imports,
- ➢ 3 lines of code.

Use PrimeQA to build your own QA app / search engine

- Head over to https://github.com/primeqa/create-primeqa-app



Conclusion

- To make information access really possible quickly we need to share code and models

- Our software needs to be compatible with one another

- PrimeQA: Let’s work on this together and make QA research move quicker than ever

PrimeQA Collaborators

	Stanford NLP		University of Illinois
	University of Stuttgart		University of Notre Dame
	Ohio State University		Carnegie Mellon University
	University of Massachusetts		

Thank you!

PrimeQA

- Clone the repo!
- Star and watch the repo
- Get regular updates
- Join the slack channel

https://github.com/primeqa/primeqa

	Stanford NLP		University of Illinois
	University of Stuttgart		University of Notre Dame
	Ohio State University		Carnegie Mellon University
	University of Massachusetts		

A Simple Assignment: Extra credits

- Get a hands-on experience working with the PrimeQA toolkit/models
- solve an open retrieval question answering task over a real world dataset: covid-qa.
- The target domain is Covid19 related documents/ publications, over which PrimeQA/models can answer natural language questions.
- We have designed experiments to focus on domain adaptation aspect of question answering.
- We will provide the Jupyter notebook – just provide the scores you get by running the models
  - Don’t forget to use GPUs ☺
- Office hours: Tuesday and Thursday (Time + Webex to be announced)