

PolQA: Polish Question Answering Dataset

Piotr Rybak, Piotr Przybyła, Maciej Ogrodniczuk

Institute of Computer Science

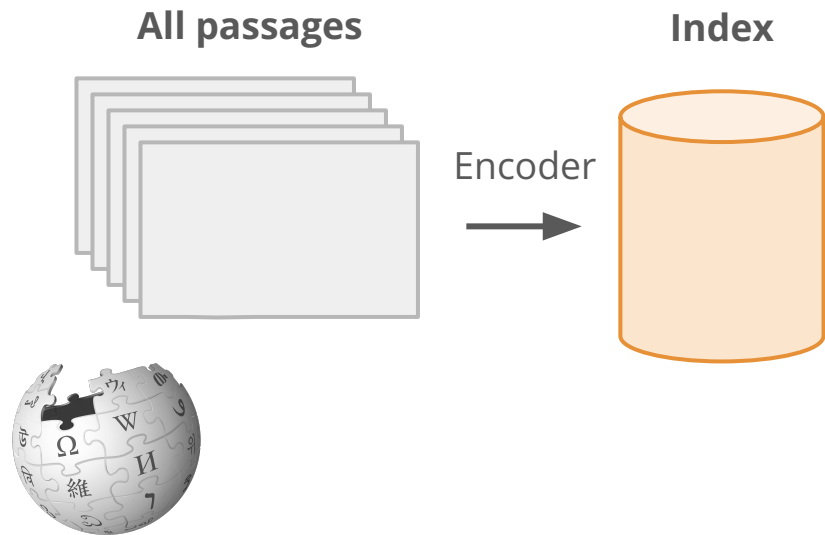
Polish Academy of Sciences

Retrieve and Read

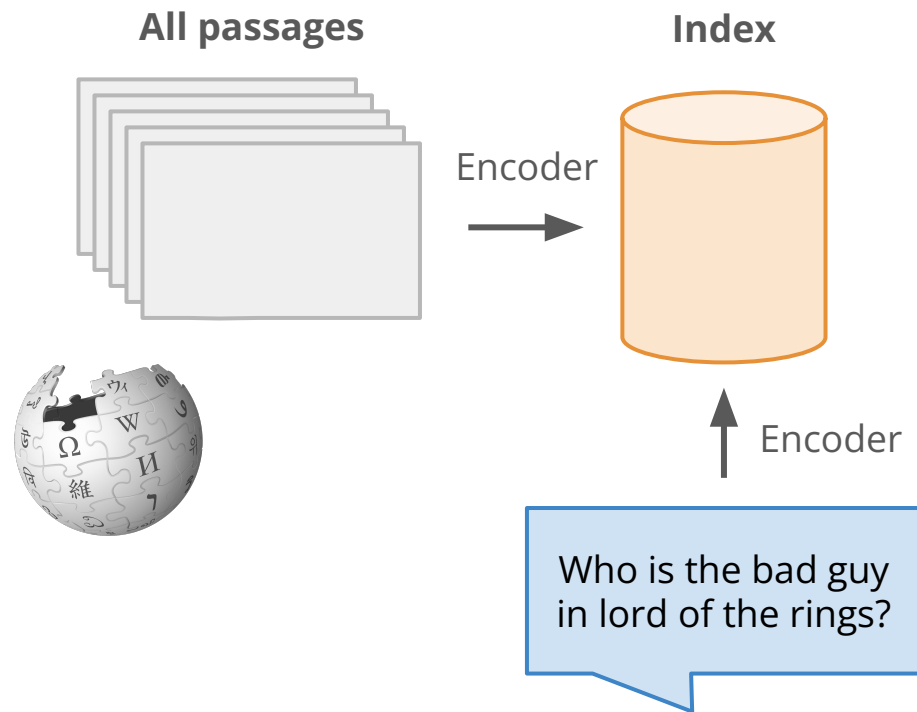
All passages



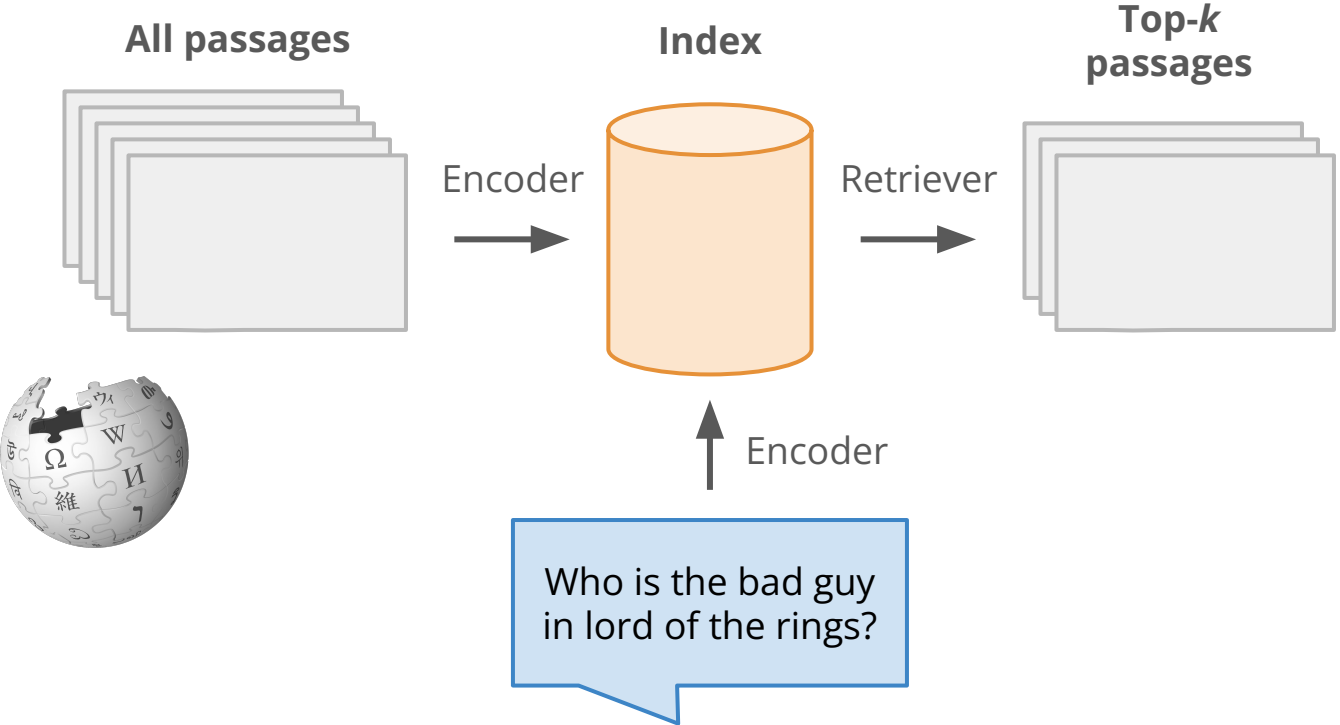
Retrieve and Read



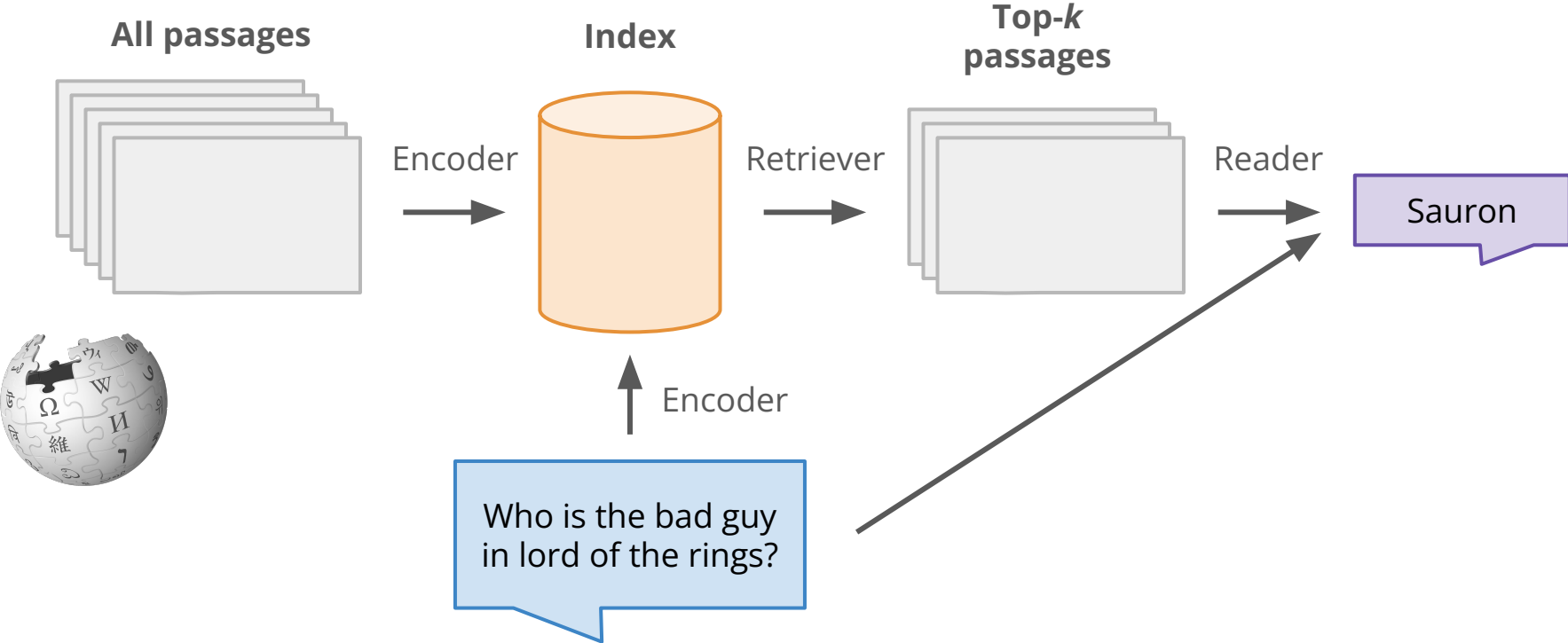
Retrieve and Read



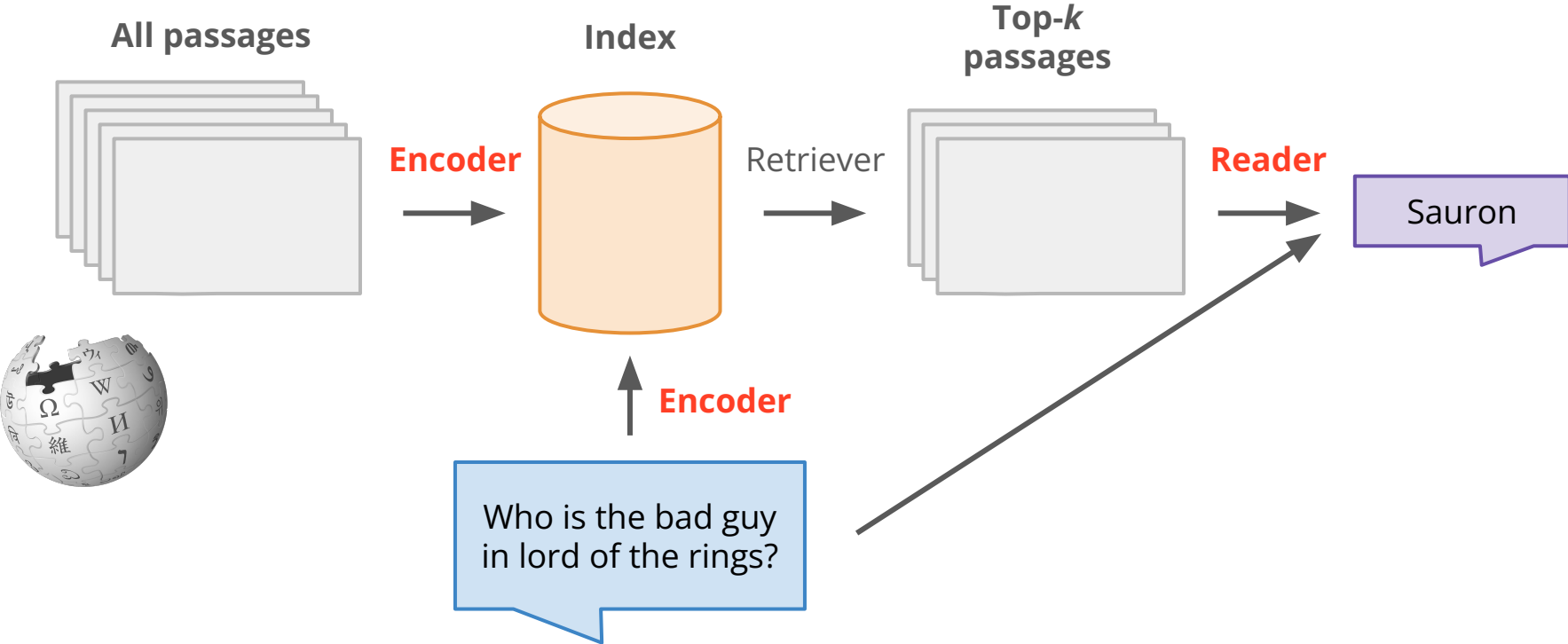
Retrieve and Read



Retrieve and Read




Retrieve and Read



PolEval [2021]

> **6 000** question-*answer* pairs but **no passages** or articles

POLEVAL 2021 Home Tasks Dates Results Prizes Publication Organizers 2020



Task 4: Question Answering Challenge

Task definition

The goal of the task is to develop a solution capable of providing answers to general-knowledge questions typical for popular TV quiz shows, such as [Fifteen to One \(PL: Jeden z dziesięciu\)](#).

The participants will be provided with:

- a small development set (question and answers)
- test set A (without answers; the results will be immediately shown in the leaderboard)
- test set B (without answers; the results will be shown in the leaderboard in the last week of the competition).

No training set will be provided this time, but participants are free to use all kinds of available offline sources. Thus, caching Wikipedia is fine but asking Google on the fly is not; the system should be ready to run without an Internet connection. The answering process also has to be completely automated and not involve any form of human assistance.

Who is the hero in the Tomb Raider video game series?

Lara Croft

Is calcium a metal?

Yes

PolQA

Who is the hero in the Tomb Raider video game series?

Lara Croft

Is calcium a metal?

Yes

PolQA

Who is the hero in the Tomb Raider video game series?

Lara Croft

Lara Croft is a character and the main protagonist of the video game franchise Tomb Raider.

Lara Croft

Is calcium a metal?

Calcium

As an alkaline earth metal calcium is a reactive metal that forms a dark oxide-nitride layer when exposed to air.

Yes

STANDARD APPROACH

> **Standard** approach:

Who is the hero in the
Tomb Raider video game series?



> **Standard** approach:

Who is the hero in the Tomb Raider video game series?



Tomb Raider

The franchise focuses on the

Tomb Raider II

It is the second entry in the

Lara Croft

Lara Croft is a character and the main protagonist of the video game franchise Tomb Raider.

It is the second entry in the Tomb Raider series and follows Lara Croft

fictional British archaeologist Lara Croft.

Strategy	# Questions			# Passages		
	Positive	Negative	Total	Positive	Negative	Total
Manual retrieval	6,296	1,763	7,000	21,451	2,729	24,180

> **Standard** approach:

Neural
retriever

> **Standard** approach:

Who is the hero in the
Tomb Raider video game series?

Neural
retriever



> **Standard** approach:

Who is the hero in the Tomb Raider video game series?

Neural retriever



- Lara Croft**
Lara Croft is a character and the main protagonist of the video game franchise Tomb Raider.
- Simon West**
West directed the 2001 action film Lara Croft: Tomb Raider.
- Tomb Raider II**
It is the second entry in the Tomb Raider series and follows Lara Croft

> **Standard** approach:

Who is the hero in the Tomb Raider video game series?

Neural retriever



Lara Croft
Lara Croft is a character and the main protagonist of the video game franchise Tomb Raider.



Simon West
West directed the 2001 action film Lara Croft: Tomb Raider.



Tomb Raider II
It is the second entry in the Tomb Raider series and follows Lara Croft



Strategy	# Questions			# Passages		
	Positive	Negative	Total	Positive	Negative	Total
Manual retrieval	6,296	1,763	7,000	21,451	2,729	24,180
Neural retrieval	4,456	6,868	7,000	8,714	26,286	35,000

EFFICIENT APPROACH

> **Efficient** approach:

Who is the hero in the Tomb Raider video game series?

Lexical &
Reranking



Lara Croft
Lara Croft is a character and the main protagonist of the video game franchise Tomb Raider.



Simon West
West directed the 2001 action film Lara Croft: Tomb Raider.



Tomb Raider II
It is the second entry in the Tomb Raider series and follows Lara Croft



Strategy	# Questions			# Passages		
	Positive	Negative	Total	Positive	Negative	Total
Standard	6,427	6,886	7,000	29,841	28,991	58,832
Manual retrieval	6,296	1,763	7,000	21,451	2,729	24,180
Neural retrieval	4,456	6,868	7,000	8,714	26,286	35,000
Efficient	5,402	6,505	7,000	13,191	21,809	35,000

Results

Model	Annotation Time [s]
Standard approach	376
Efficient approach	68

Results

Model	Annotation Time [s]	Accuracy@10
Standard approach	376	51.5%
Efficient approach	68	62.0%

Evaluated on a test set gathered using Standard approach

PolQA

> **7 000** questions

> **39k** positive pairs

> **48k** hard negative pairs

Model	Annotation Time [s]	Accuracy@10
Standard approach	376	51.5%
Efficient approach	68	62.0%

Evaluated on a test set gathered using Standard approach

THANK YOU

piotr.cezary.rybak@gmail.com