

PolEval 2024 Task 1: Reading Comprehension

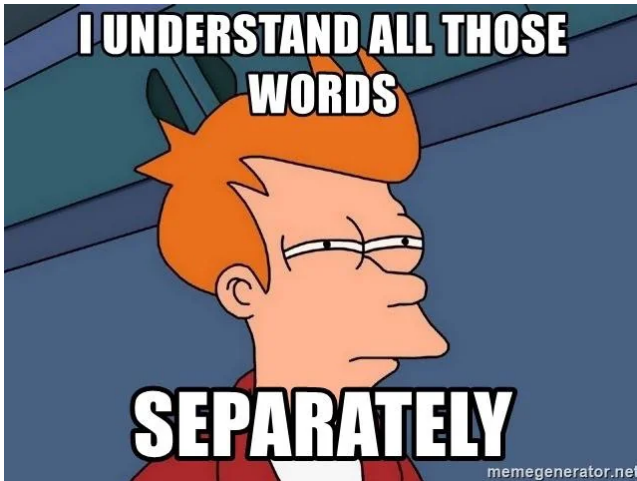
Ryszard Tuora, Aleksandra Zwierzchowska

Institute of Computer Science, Polish Academy of Sciences

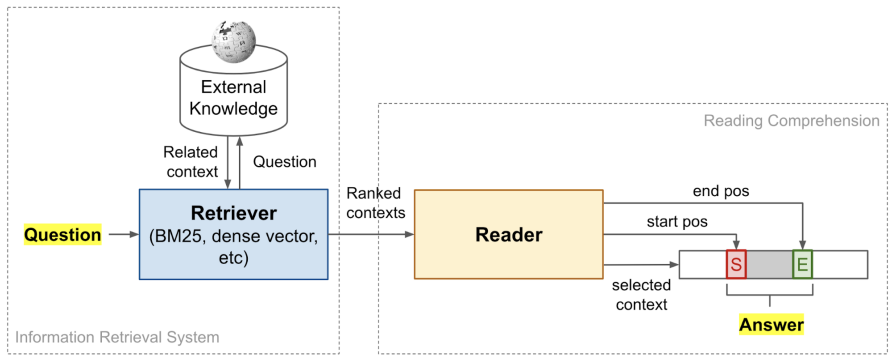
December 2, 2024

- 1 Introduction
- 2 Task definition
- 3 Dataset
- 4 Evaluation
- 5 Baseline
- 6 Submission and results

Reading Comprehension



Classical Question Answering System



Span Extraction Limitations

- Does not work well for morphologically rich languages
- Does not fit well with answering yes-no questions
- Limits the complexity of reasoning that might be used to answer the question

Generative Question Answering

Question

Kto wygrał turniej w Montrealu?

Passage

(...) w trzeciej rundzie przegrał z późniejszym triumfатorem turnieju **Rafaelem Nadalem**.

Answer

Rafael Nadal

Data provided

- **Training set**

- Paragraphs: 11 624
- Questions: 56 618

- **Dev set**

- Paragraphs: 1 453
- Questions: 7 060

PoQuAD



Polish Question Answering Dataset (Tuora et al. 2022, 2023)

- Part of the CLARIN-PL-Biz project, conducted at Wrocław University of Science and Technology
- Polish Wikipedia
- Generative answer layer
- Inclusion of impossible questions

Extractive and Generative answer layer

Type of change	Extractive	Generative
inflection	drugorzędny śródoceaniczny	drugorzędnym śródoceanicznym
grammatical class	podpalić meczet	podpalenia meczetu
adding a word	czasów saskich	w czasach saskich
eliminating words	owłosieniem liści i załąźni, cechuje się także innym zasięgiem występowania	owłosieniem liści i załąźni, zasięgiem występowania
yes/no	podał się do dymisji	tak

Impossible questions



Impossible questions

Question

Jaki obiekt wybudowany przez Wybickiego zachował się w Manieczkach do dziś?

Passage

Obecnie jedyną pozostałością z czasów Józefa Wybickiego jest niewielka **barokowa kaplica** w kształcie rotundy, ufundowana przez Wybickiego i wybudowana w 1786.

Metrics

- **Normalized Levenshtein Distance**
- **Binary F1 answerability classification**

Levenshtein Distance (Edit Distance)

Minimum number of single-character edits (insertions, deletions or substitutions) required to change one string into the other

Examples

Rafael Nadal ↔ Rafaelem Nadalem ⇒ **4**

wróble zwyczajne i mazurki ↔ wróble zwyczajne ⇒ **10**



Power of the system to answer questions

F1 score

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

$$\textit{Precision} = \frac{TP}{TP + FP}$$

How many retrieved items are relevant?

$$\textit{Recall} = \frac{TP}{TP + FN}$$

How many relevant items are retrieved?



Capacity to recognize when the information is sufficient,
and when to abstain from answering

Baseline

	Normalized Levenshtein	F1	Score
GPT 3.5 few shot	67.25	48.20	57.73
pIT5 baseline	83.25	57.67	70.46

*GPT 3.5 few shot - paragraph from the development set listed as example

Submission and results

	Normalized Levenshtein	F1	Score
Krzysztof Wróbel	84.78	81.89	83.34

Congratulations to the winner!