

Passage Retrieval in question answering systems in Polish language

Anna Pacanowska

Instytut Informatyki UWr

19 września 2023

My models

Different approaches:

- Lemmatization
- Transformer-based models

Lemmatization

Definition (Lemmatization)

Process of transforming a word into its dictionary form ('lemma').

'byli' -> 'być'

'koty' -> 'kot'

'zrobiony' -> 'zrobić'

Observation

Lemma depends on the context.

'To w złym tonie: laptop tonie w tonie betonu.' -> 'To w złym ton: laptop tonąc w tona beton.'

BM25 vs lemmatization

1 Unlemmatized

- ▶ query: 'How long will I wait for my letter?'
- ▶ passage: 'Letters' waiting time is usually around 3 days.'
- ▶ matches: 0

2 Lemmatized

- ▶ query: 'How long will I wait for my letter?'
- ▶ passage 'Letter wait time be usually around 3 day.'
- ▶ matches: 2 (letter, wait)

Lemmatization models

- Morfeusz2
 - ▶ dictionary based
 - ▶ provides all possible lemmas
- spaCy
 - ▶ neural network based
 - ▶ tries to choose the correct lemma

Lemmatization methods

① Morfeusz2

problems:

- ▶ multiple possibilities
- ▶ false matches

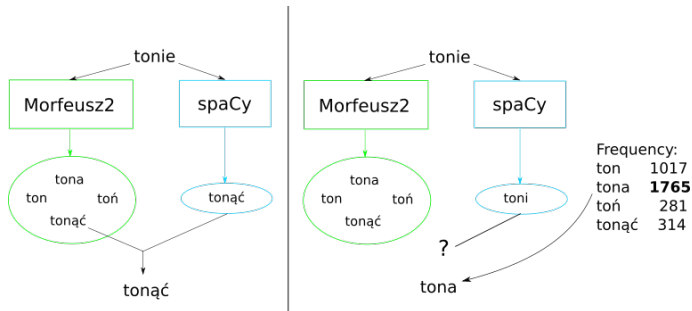
② spaCy

problems:

- ▶ incorrect lemmas
- ▶ invalid lemmas
- ▶ slower

Lemmatization methods

- 3 Morfeusz2 + spaCy
- 4 Morfeusz2 + pick the most popular lemma



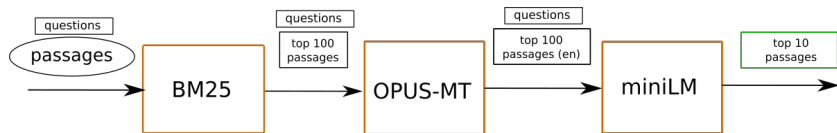
Lemmatization results

method	NDCG@10
no lemmatization	18.62
spaCy	21.41
Morfeusz2	21.15
hybrid	24.47
Morfeusz2-freq	25.24

Different lemmatization methods (dev dataset)

Transformer-based model: basic outline

- 1 Pick 100 candidates using BM25 (with Morfeusz2 lemmatization).
- 2 Translate questions and passages (OPUS-MT).
- 3 Re-rank using pre-trained model (miniLM).



Why miniLM?

- Performs very well on BEIR benchmark
- Small and fast

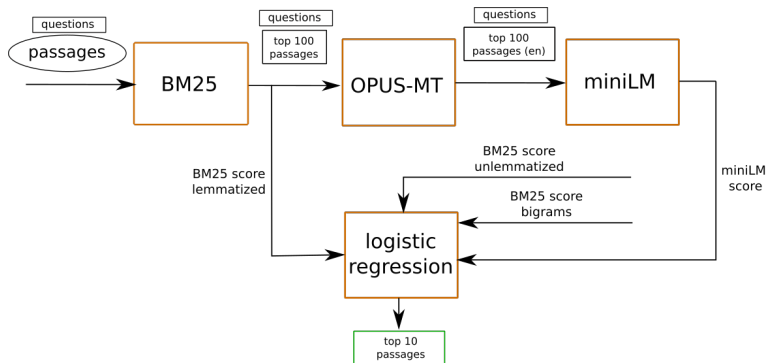
Basic approach – results

model	dev	test-A
Baseline (from PolEval)	-	50.76
Morfeusz2-freq	25.24	-
miniLM	31.36	58.19

Joining different scores

Train logistic regression on:

- miniLM score
- unlemmatized BM25
- lemmatized BM25
- bigrams BM25

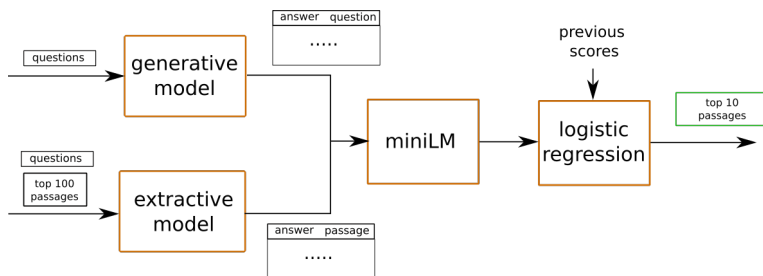


Joining different scores – results

model	test-A
miniLM	58.19
miniLM with logistic regression	60.17

Add answers from QA models

- Two QA models
 - ▶ Generative: add answer to question (GPT3, chatGPT)
 - ▶ Extractive: add answer to passage (DistilBERT)
- Similar answers -> likely relevant passage
- Useful for factoid, common knowledge questions



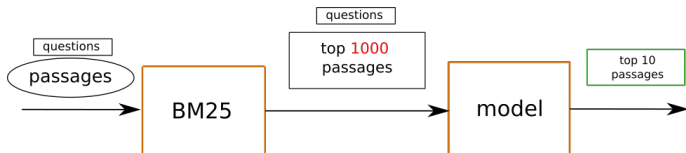
Add answers from QA models – results

model	test-A
miniLM	58.19
miniLM with logistic regression	60.17
miniLM with logistic regression, GPT3	60.97
miniLM with logistic regression, GPT3, chatGPT	56.18

More passages to re-rank

Relevant passages in BM25-Morfeusz ranking (train set):

:100	101:1000	1001:
6726	3044	4678



More passages to re-rank – results

model	test-A
miniLM	58.19
miniLM with logistic regression	60.17
miniLM with logistic regression, GPT3	60.97
miniLM with logistic regression, GPT3, chatGPT	56.18
miniLM with logistic regression, GPT3, 1000 passages	62.51

Domains

model	test-A			test-B		
	wiki	legal	allegro	wiki	legal	allegro
BM25	19.76	81.10	49.16	18.45	80.32	48.05
my best model	38.27	77.70	69.96	37.11	79.00	67.92

Scores on different domains.

Acknowledgements

This paper is taken from my master's thesis written under the direction of dr Paweł Rychlikowski at the University of Wrocław.

Thank you for attention!