Answering Polish Trivia Questions with the Help of Dense Passage Retriever

Aleksander Smywiński-Pohl, Dmytro Zhylko,
Krzysztof Wróbel, Magdalena Król

AGH, UJ, Enelpol

Tested Approaches

Tested Approaches

- selective QA

Tested Approaches

- selective QA
- extractive QA

Tested Approaches

- selective QA
- extractive QA
- closed-book QA

Selective QA

Selective QA

- "Czy wiesz" dataset
- LQuAD-PL
- distant supervision

Selective QA

- "Czy wiesz" dataset
- LQuAD-PL
- distant supervision

Dense passage retriever

Distant supervision

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)
- retriever first trained on "Czy wiesz" and LQuAD-PL

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)
- retriever first trained on "Czy wiesz" and LQuAD-PL
- top-100 passages

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)
- retriever first trained on "Czy wiesz" and LQuAD-PL
- top-100 passages
- select first passage containing answer as **Positive**

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)
- retriever first trained on "Czy wiesz" and LQuAD-PL
- top-100 passages
- select first passage containing answer as **Positive**
- select remaining top passages as **Negative**

Distant supervision

- divide Wikipedia into passages (90 words, full sentence)
- retriever first trained on "Czy wiesz" and LQuAD-PL
- top-100 passages
- select first passage containing answer as **Positive**
- select remaining top passages as **Negative**
- train new retriever on extended dataset (Czy wiesz, LQuAD-PL + Task-4)

Dense Passage Retriever

Dense Passage Retriever

Training:

- 20 epochs
- 3,7 million passages
- best epoch 15/16

Dense Passage Retriever

Training:

- 20 epochs
- 3,7 million passages
- best epoch 15/16

Top-1 accurracy@dev

Herbert-base 63,6%

Herbert-large 68,0%

Ranker + Reader - DPR

Ranker + Reader - DPR

- re-ranking top 80 passages

Ranker + Reader - DPR

- re-ranking top 80 passages
- top-1 is selected

Ranker + Reader - DPR

- re-ranking top 80 passages
- top-1 is selected
- extractive QA

Ranker + Reader - DPR

- re-ranking top 80 passages
- top-1 is selected
- extractive QA

Retriever + reader: 46.84%

Boolean questions

Boolean questions

- Selection "Czy" but not "czy" :-)

Boolean questions

- Selection "Czy" but not "czy" :-)
- top-1 is selected

Boolean questions

- Selection "Czy" but not "czy" :-)
- top-1 is selected
- NLI - CDSC as similar task

Boolean questions

- Selection "Czy" but not "czy" :-)
- top-1 is selected
- NLI - CDSC as similar task
- yes - entailment, no - contradiction

Boolean questions

- Selection "Czy" but not "czy" :-)
- top-1 is selected
- NLI - CDSC as similar task
- yes - entailment, no - contradiction

Retriever + reader + yes/no: 50.96%

Closed-book QA

Closed-book QA

- mT5 model family
- pure text-to-text

Closed-book QA

- mT5 model family
- pure text-to-text

Best solution

Best solution

- retriever HerBERT-base

Best solution

- retriever HerBERT-base
- top-80 results

Best solution

- retriever HerBERT-base
- top-80 results
- reader HerBERT-large

Best solution

- retriever HerBERT-base
- top-80 results
- reader HerBERT-large
- 12 negative contexts

Best solution

- retriever HerBERT-base
- top-80 results
- reader HerBERT-large
- 12 negative contexts
- batch size: 1

Best solution

- retriever HerBERT-base
- top-80 results
- reader HerBERT-large
- 12 negative contexts
- batch size: 1
- NLI: HerBERT-large

Key takeaways

Key takeaways

- dev dataset has to resemble **test** rather than train dataset
  - 24% → 46%

Key takeaways

- dev dataset has to resemble **test** rather than train dataset
  - 24% → 46%
- model size has a huge impact
  - base → large: 49% → 60%

Key takeaways

- dev dataset has to resemble **test** rather than train dataset
  - 24% → 46%
- model size has a huge impact
  - base → large: 49% → 60%
- number of negative samples (re-ranker) is important
  - 3 →10: 21% →32%
  - 10 →15: 32% →49%

Thank you!

Demo in the legal domain

https://lemkin.pl