

Simple Recipes for Question Answering

Darek Kłeczek

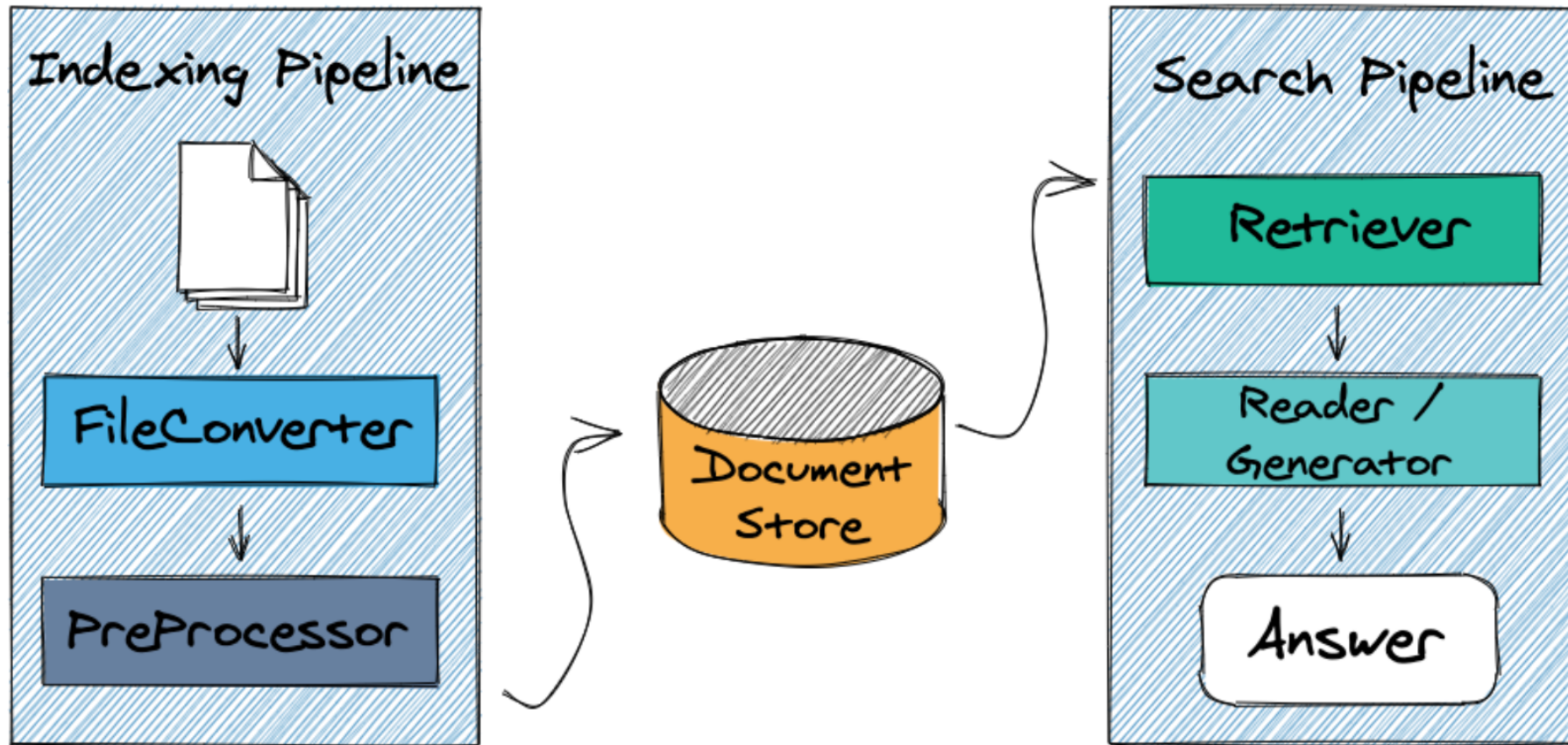
Closed Book Approach



- PapuGaPT2
- Finetune on dev set:
Pytanie: <Q> Odpowiedź: <A> <EOS>
- 18% accuracy

<https://huggingface.co/flax-community/papuGaPT2>

Open Book Approach



Reader

- mBERT-base pretrained by Henryk Borzymowski on translated SQUAD
- XLM-Roberta-large trained on English SQUAD (+ **10p**)

Retriever

- BM25
- DPR – failed attempt (using distant supervision)

Postprocessing

- Yes/No -> Yes
- Numbers -> numeric format
- Multiple choice -> prediction if matching or last option

Data-driven Improvements

- Polish Wikipedia dump
- Polish word definitions from Polish Wordnet
- Polish proverbs
- Crawl: brainly, sciaga, bryk.pl, opracowania (via mc4)

Thank You