

Task 3: Passage Retrieval

Piotr Rybak

(Instytut Podstaw Informatyki PAN)

Aleksander Smywiński-Pohl

(Akademia Górniczo-Hutnicza)

Piotr Przybyła

(IPI PAN / Uniwersytet Pompeu Fabra)

Maciej Ogrodniczuk

(Instytut Podstaw Informatyki PAN)

Motivation

How to find passages containing answer to the given questions?

- Searching using keywords works surprisingly well, but:
 - requires exact match of keywords between the question and a passage,
 - ignores word order.
- Neural models works better when trained on the in-domain data, but:
 - requires a dataset for training,
 - doesn't generalize well to different domains.

Task description

Create a cross-domain retrieval system:

- Given the question find relevant passages in a given corpus,
- Training set from a single domain (trivia),
- Test sets from three different domains (trivia, legal, customer support).



Training data

5 000 trivia questions, 16 000 passages

W którym roku
wynaleziono tranzystor?

Pierwszy działający tranzystor ostrykowy został
skonstruowany 16 grudnia 1947 r.

Test data

1 300 trivia questions, 700 legal questions, 900 support questions

Ile budżetów ma miasto na prawach powiatu?

Art. 85. 1. Miasto na prawach powiatu sporządza jeden budżet. 2. Uchwała budżetowa miasta na prawach powiatu określa ...

Czy mogę zmienić swoją nazwę użytkownika?

Niestety, ze względów bezpieczeństwa, zmiana raz nadanej nazwy użytkownika (loginu) nie jest możliwa.

Evaluation & Results

Eval metric: Normalized Discounted Cumulative Gain at 10

Submission	Test-B NDCG@10
Jakub Pokrywka	69.36
Marek Kozłowski	68.19
Konrad Wojtasik	67.44
Norbert Ropiak	63.27
Anna Pacanowska	54.23
Maciej Kazuła	51.78
Daniel Karaś	51.71

Thank you!

