MULTIMODAL DOCUMENT UNDERSTANDING THROUGH VISUAL QUESTION ANSWERING

Šárka Ščavnická

WHICH MODELS

- LayoutXLM
- LayoutLMv2 base
- LayoutLMv3 base
- Layoutlm-invoices from impira
  - Fine-tuned on a proprietary dataset of invoices
  - Fine-tuned on SQuAD2.0 and DocVQA for general comprehension.

METRICES

- **Based on both start and end position**
  - f1, precision, accuracy
- **Just start position:**
  - Recall, f1, precision
- **Just end position**
  - Recall, f1, precision
- **Based on the textual answer**
  - Recall, f1, precision

GRAPHS

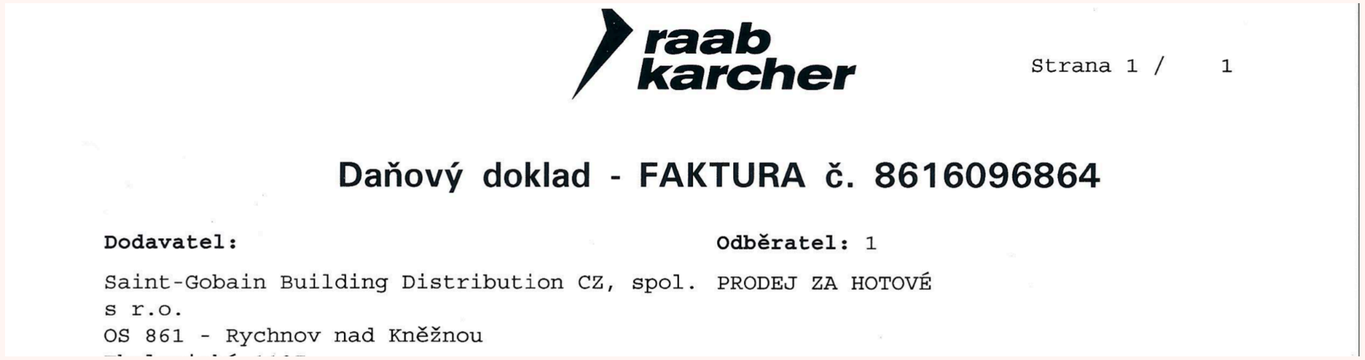
- https://wandb.ai/diplomka_dvqa/DVQA-CZ_001_v3_metrics?workspace=user-scavnicka-sarka

USING TRAINED MODEL

- **Script test.py**

```
Číslo faktúry?

Predicted start idx: 28
Predicted end idx: 31
8616096864
```



raab karcher

Strana 1 / 1

Daňový doklad - FAKTURA č. 8616096864

Dodavatel:

Saint-Gobain Building Distribution CZ, spol. s r.o.
OS 861 - Rychnov nad Kněžnou

Odběratel: 1

PRODEJ ZA HOTOVÉ

- **Q_invoice_number = ['Jaké je číslo faktury?', 'Pod jakým číslem je vedena faktura?', 'Číslo faktury?', 'Jaké je označení faktury?']**

```
Faktura č.?

Predicted start idx: 27
Predicted end idx: 30
8616096864
```

BRAINSTORMING

- New entities?
  - Can not do it?
  - Trained on smaller sets (10 examples) for each new entity?