MUNI
FI

Towards General Document Understanding through QA

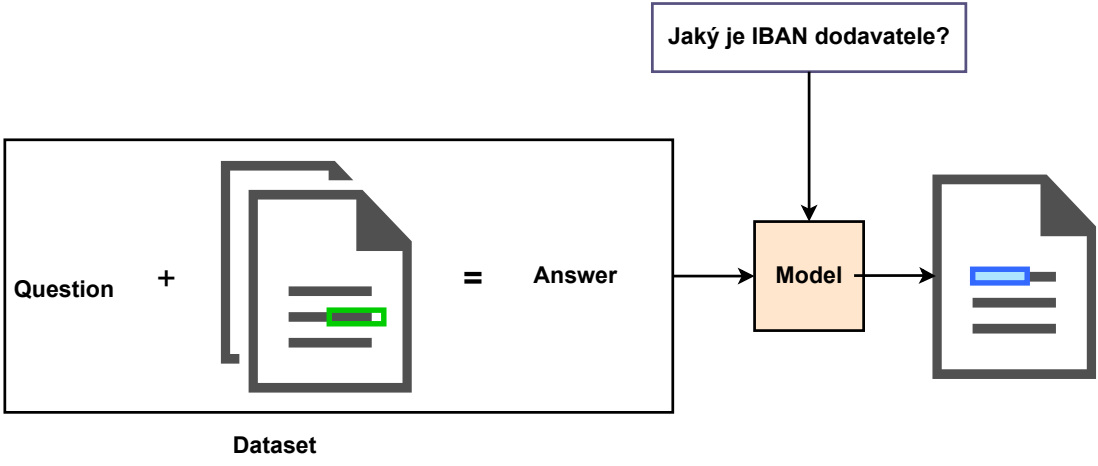
***Šárka Ščavnická*, Michal Štefánik, Marek Kadlčík,**
Martin Geletka, Petr Sojka
527352@mail.muni.cz, sojka@fi.muni.cz

Math Information Retrieval Research Group,
Faculty of Informatics, Masaryk University

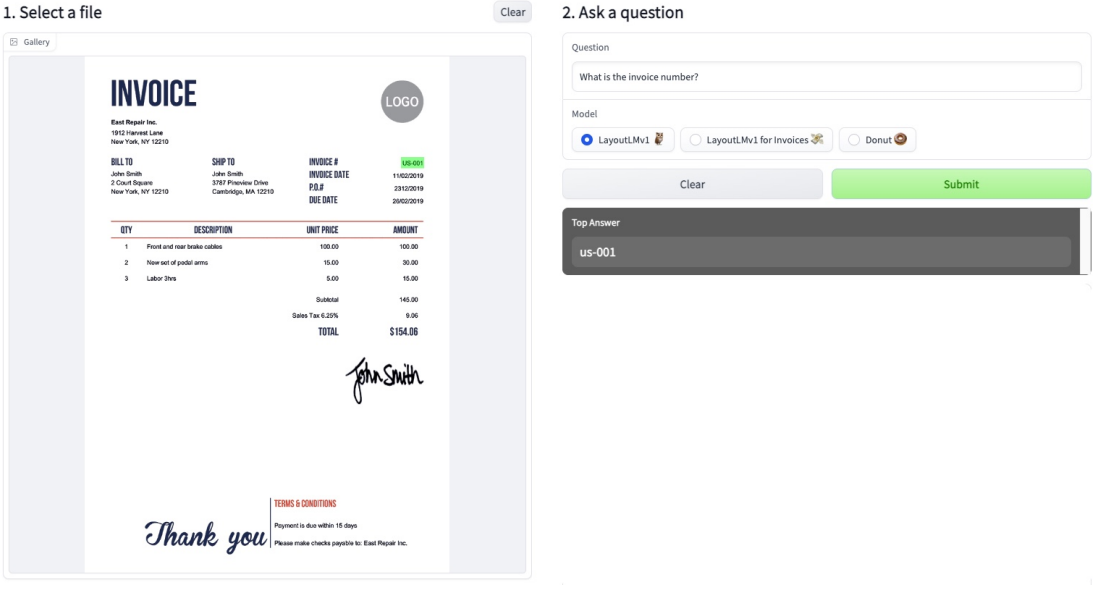
December 9th, 2022

Motivation

Model Description



Competitive system for the English language



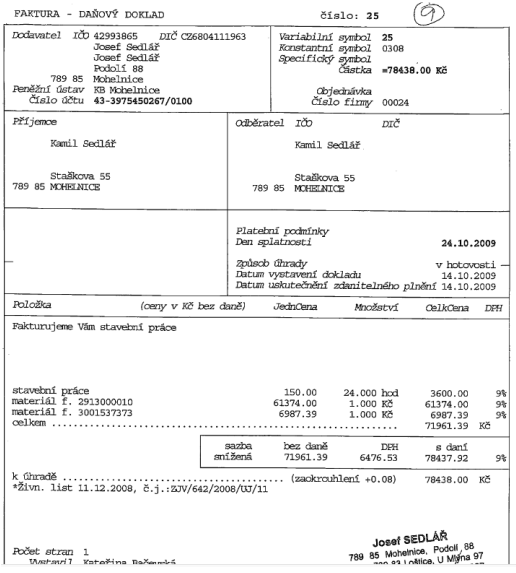
Textual Representation of the Invoice

- FAKTURA - DAŇOVÝ DOKLAD © číslo: (25 Dodavatel | IČO 42993865 DIČ |CZ6804111963| Variabilní symbol 25 Josef Sedlář Konstantní symbol [0308 Josef Sedlář Specifický symbol , Podolí 88 Částka 78438.00 Kč 789 85 Mohelnice Peněžní ústav KB Mohelnice Objednávka Číslo účtu (100 Číslo fixmy 00024 Příjemce Odběratel IČO Drč Kamil Sedlář Kamil Sedlář Staškova 55 Staškova 55 789 85 MOHELNICE 789 85. MOHELNICE Platební Den splatnosti 24.10.2009
  - Who is the supplier, and what is his address?

Textual Representation of the Invoice

- FAKTURA - DAŇOVÝ DOKLAD © číslo: (25 **Dodavatel** | IČO 42993865 DIČ |CZ6804111963| Variabilní symbol 25 Josef Sedlář Konstantní symbol [0308 Josef Sedlář Specifický symbol , Podolí 88 Částka 78438.00 Kč 789 85 Mohelnice Peněžní ústav KB Mohelnice Objednávka Číslo účtu (100 Číslo fixmy 00024 Příjemce Odběratel IČO Drč Kamil Sedlář Kamil Sedlář Staškova 55 Staškova 55 789 85 MOHELNICE 789 85. MOHELNICE Platební Den splatnosti 24.10.2009
  - Who is the supplier, and what is his address?
- We need more context

Visual Representation of the Invoice



FAKTURA - DAŇOVÝ DOKLAD Číslo: 25

Dodavatel IČO 42993865 DIČ CZ6804111963
Josef Sedlář
Josef Sedlář
Podolí 88
789 85 Mohelnice
Peněžní ústav KB Mohelnice
Číslo účtu 43-3975450267/0100

Variabilní symbol 25
Konstantní symbol 0308
Specifický symbol
Částka =78438.00 Kč
Objednávka
Číslo firmy 00024

Příjemce
Kamil Sedlář
Staškova 55
789 85 MOHELNICE

Odběratel IČO DIČ
Kamil Sedlář
Staškova 55
789 85 MOHELNICE

Platební podmínky
Den splatnosti 24.10.2009
Způsob úhrady v hotovosti
Datum vystavení dokladu: 14.10.2009
Datum uskutečnění zdanitelného plnění 14.10.2009

Položka (ceny v Kč bez daně)	JednCena	Množství	CelkCena	DPH
Fakturujeme Vám stavební práce				
stavební práce	150.00	24.000 hod	3600.00	9%
materiál f. 2913000010	61374.00	1.000 Kč	61374.00	9%
materiál f. 3001537373	6987.39	1.000 Kč	6987.39	9%
celkem			71961.39 Kč	

sazba	bez daně	DPH	s daní	
snížená	71961.39	6476.53	78437.92	9%

k úhradě ..... (zaokrouhlení +0.08) 78438.00 Kč
*Živn. list 11.12.2008, č.j.:ŽÚV/642/2008/ÚÚ/11

Počet stran 1
Vystavil Kateřina Bačkovská

Josef SEDLÁŘ
789 85 Mohelnice, Podolí 88

Text Representation of the Invoice with the Visual Kknowledge

- Who is the supplier, and what is his address?
- FAKTURA - DAŇOVÝ DOKLAD © číslo: (25 **Dodavatel** | IČO 42993865 DIČ |CZ6804111963| Variabilní symbol 25 **Josef Sedlář** Konstantní symbol [0308 **Josef Sedlář** Specifický symbol , **Podolí 88** Částka 78438.00 Kč **789 85 Mohelnice** Peněžní ústav KB Mohelnice Objednávka Číslo účtu (100 Číslo fixmy 00024 Příjemce Odběratel IČO Drč Kamil Sedlář Kamil Sedlář Staškova 55 Staškova 55 789 85 MOHELNICE 789 85. MOHELNICE Platební Den splatnosti 24.10.2009
- Better results if we use the visual layout of the document

Methods and Datasets

Question Answering

- Two types:
  - Extractive QA
  - Generative QA
- Multilingual datasets, primarily in English
- Datasets:
  - Stanford Question Answering Dataset (SQuAD)
    - Wikipedia articles
  - French Question Answering Dataset (FQuAD
    - Wikipedia articles

Visual Question Answering

- Focusing only on the visual part of an image
- We can use the same images for different languages
- Datasets:
  - Visual Question Answering (VQA) dataset
  - The Image-Set Visual Question Answering (ISVQA)

Named Entity Recognition

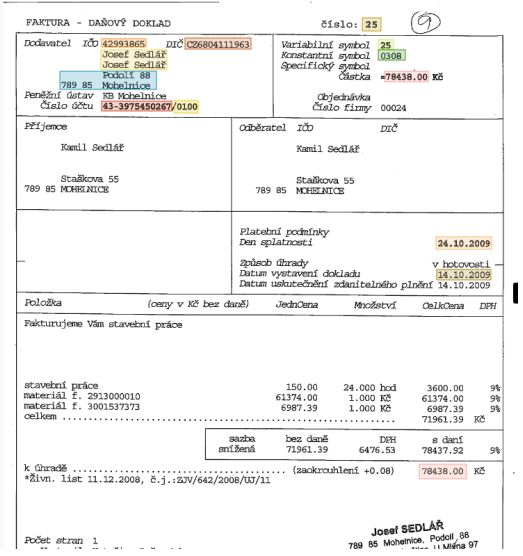
- Aims to locate and identify entities in the given text
- Datasets:
  - CoNLL-2003 dataset
    - English and German
    - Newspaper articles
  - Few-NERD
    - Wikipedia articles
  - Dataset of German Legal Documents for NER
    - Non-English
    - Legal documents

Document Visual Question Answering

- Combines the visual and textual sides of a document
- Lacking coverage of non-English models for Document Visual Question Answering
- Datasets:
  - DocVQA
    - From the UCSF Industry Documents Library
    - Contains invoices
    - Documents are from the period between 1960 and 2000
  - Visual Machine Reading Comprehension (VisualMRC)
    - Consisting of screenshots of web pages

Dataset creation

Invoices - 6,849 documents and 15 entities



FAKTURA - DAŇOVÝ DOKLAD

Číslo: 25

Dodavatel IČO 42993865 DIČ CZ6804111963
Josef Sedlář
Josef Sedlář
Podolí 88
789 85 Mohelnice
Peněžní ústav KB Mohelnice
Číslo účtu 43-3975450267/0100

Variabilní symbol 25
Konstantní symbol 0308
Specifický symbol
Částka =78438.00 Kč
Objednávka
Číslo firmy 00024

Příjemce
Kamil Sedlář
Staškova 55
789 85 MOHELNICE

Odběratel IČO DIČ
Kamil Sedlář
Staškova 55
789 85 MOHELNICE

Platební podmínky
Den splatnosti 24.10.2009
Způsob úhrady v hotovosti
Datum vystavení dokladu 14.10.2009
Datum uskutečnění zdanitelného plnění 14.10.2009

Položka (ceny v Kč bez daně)	JednCena	Množství	CelkCena	DPH
Fakturujeme Vám stavební práce				
stavební práce	150.00	24.000 hod	3600.00	9%
materiál f. 2913000010	61374.00	1.000 Kč	61374.00	9%
materiál f. 3001537373	6987.39	1.000 Kč	6987.39	9%
celkem			71961.39	Kč

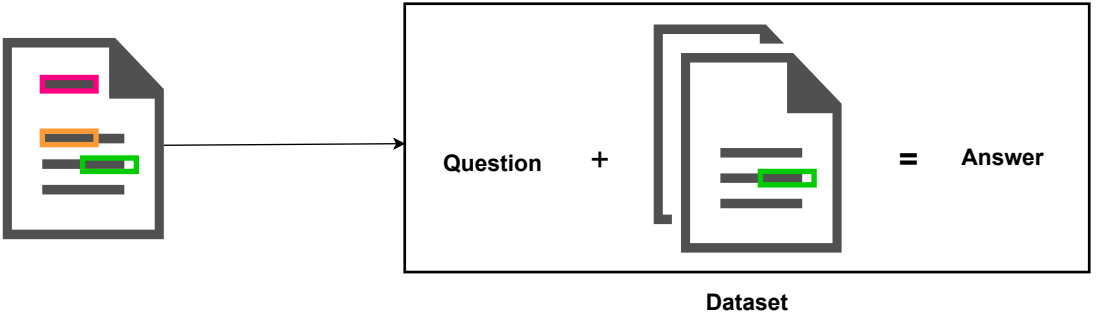
	sazba	bez daně	DPH	s daní	
	snížená	71961.39	6476.53	78437.92	9%

k úhradě ..... (zaokrouhlení +0.08) 78438.00 Kč
*živn. list 11.12.2008, č.j.:ŽÚ/642/2008/ÚJ/11

Počet stran 1

Josef SEDLÁŘ
789 85 Mohelnice, Podolí 88

Dataset Description



Research Proposal

How can Textual QA and Visual QA datasets improve Document Understanding?

First Question

- Our model will be trained on the Czech Document VQA dataset
  - A large number of relevant but narrow questions
- Hypothesis:
  - If we first introduce our model to much more general and comprehensive QA datasets
    - More robust in the unseen evaluation scenarios
    - Perform better

How well can Document VQA models generalize beyond training entity types?

Second Question

- Conventional Named Entity Recognition models
  - A closed set of entity types present in the training set
  - Must be retrained when a new entity needs to be recognized
- Document VQA
  - A closed set of questions
  - Retraining?
    - How well our model can generalize to unseen entities?

How much can Document VQA in non-English languages benefit from English datasets?

Third Question

- Compare the model performance on a target language
- Train on:
  - Only English data
  - A mixture of English and the target-language data
  - Only the target-language data
- The evaluation of the approach will assess the applicability of our models to unseen language

MUNI
FACULTY
OF INFORMATICS