

Automated Classification of Czech Municipal Council Agenda Items

Czech municipalities publish thousands of council agenda items per year in heterogeneous PDFs across city portals, with no thematic labels. I present an end-to-end pipeline combining city-specific scrapers, a schema-driven PDF parser, and a fine-tuned **RobeCzech** multi-label classifier, trained on **2,113** manually annotated items in **17** thematic categories. Test micro F1 = **0.900**.

Motivation

Agenda items cover land transfers, budget amendments, and infrastructure projects. Yet all are published as **unstructured PDFs** scattered across city portals, with **no thematic labels**. Without classification, there is **no systematic way** to search or filter council decisions by topic.

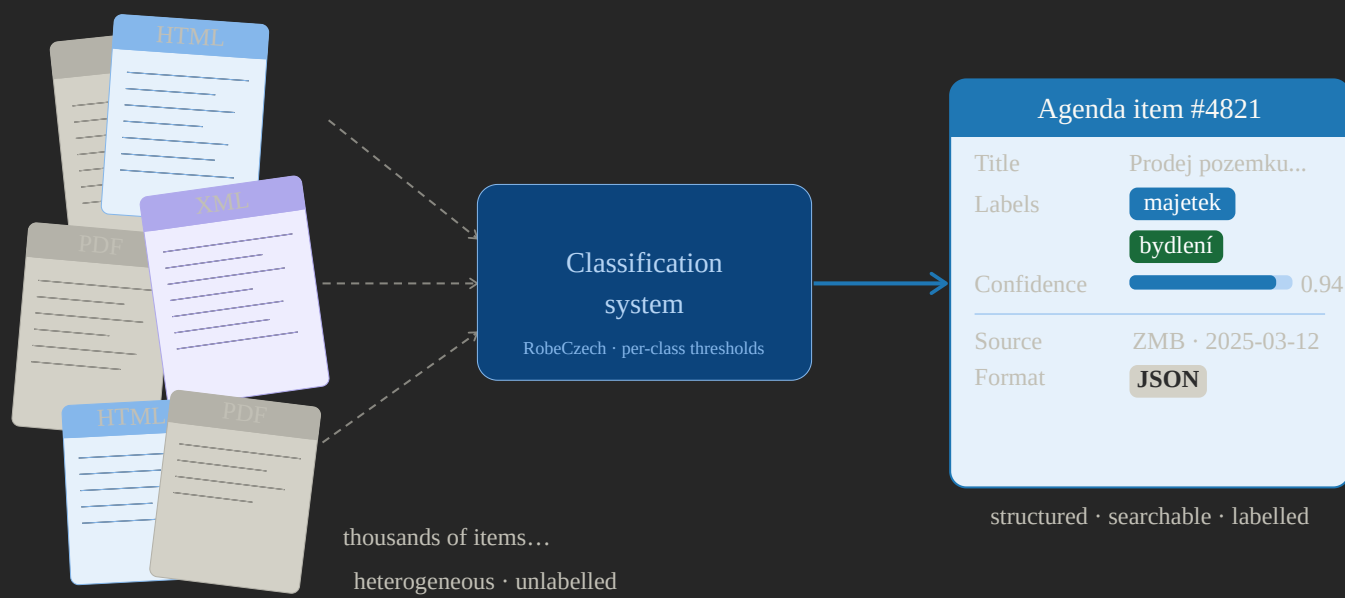


Figure 1. Heterogeneous document formats across Czech municipal portals.

Pipeline

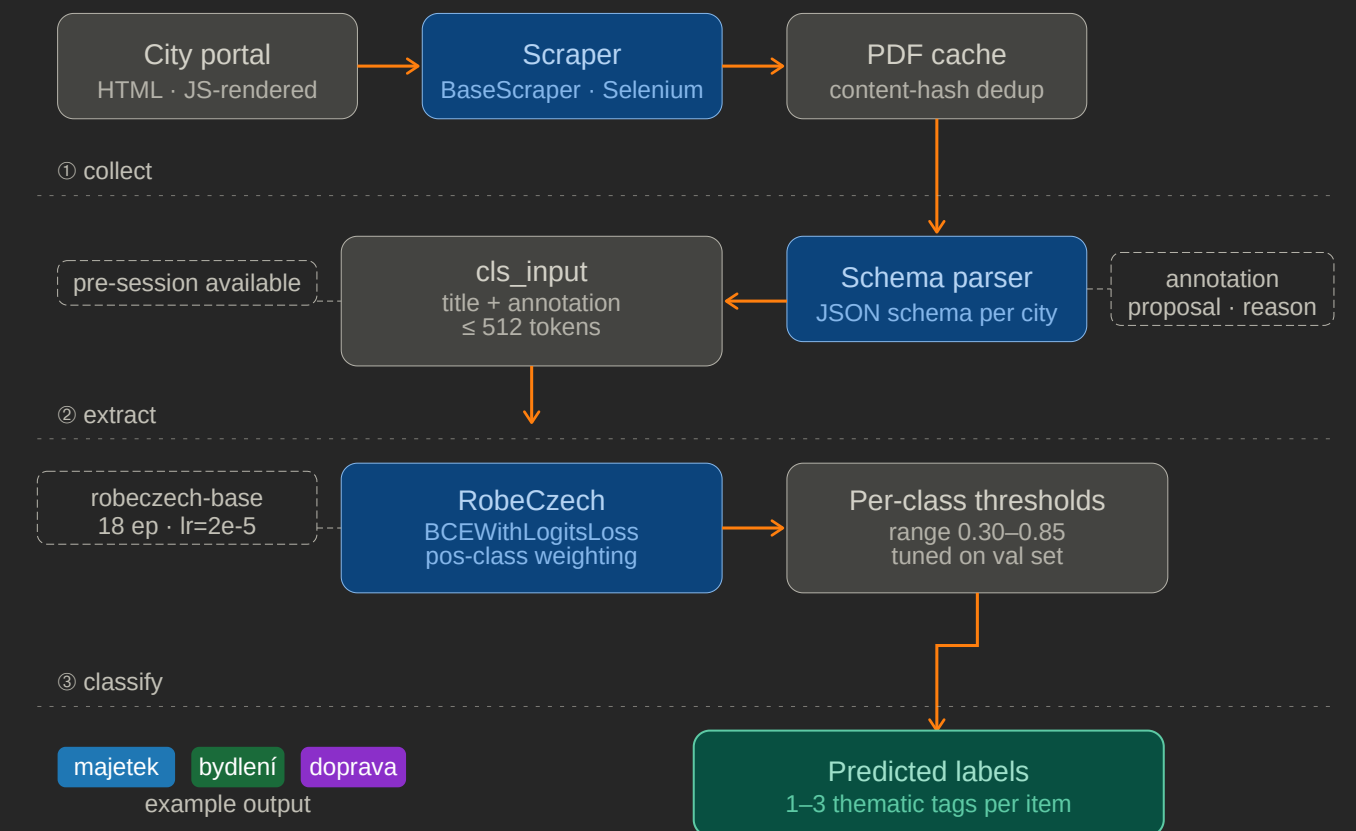


Figure 4. End-to-end pipeline.

Dataset & Taxonomy

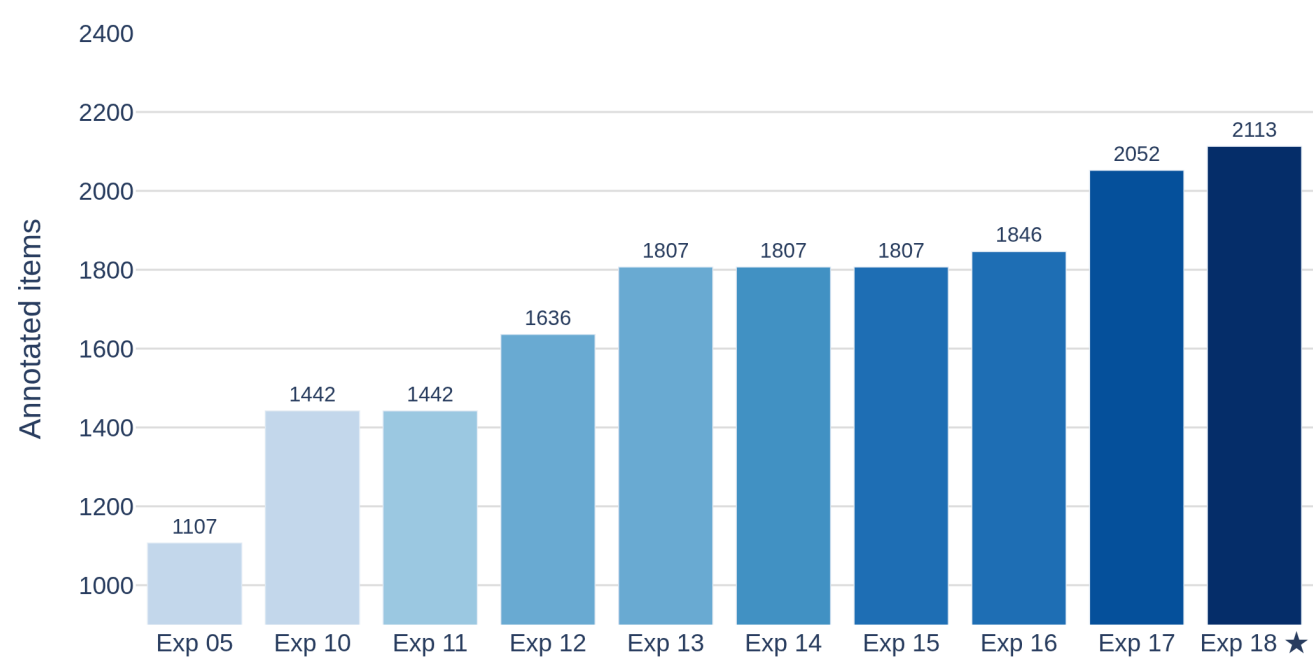
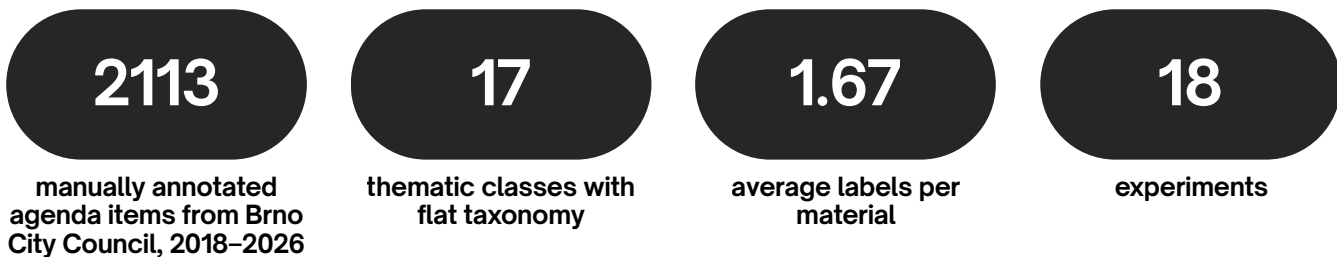


Figure 2. Dataset growth over 18 annotation iterations (Brno, 2018–2026).

Model & Training

Model: RobeCzech (ufal/rob Czech), 4.9B Czech tokens
Best config: 18 epochs · lr = 2e-5 · batch size 16
Task formulation: linear head over [CLS] token, independent sigmoid per class
Loss: BCEWithLogitsLoss with positive-class weighting

0.900

test micro-F1

0.884

test macro-F1

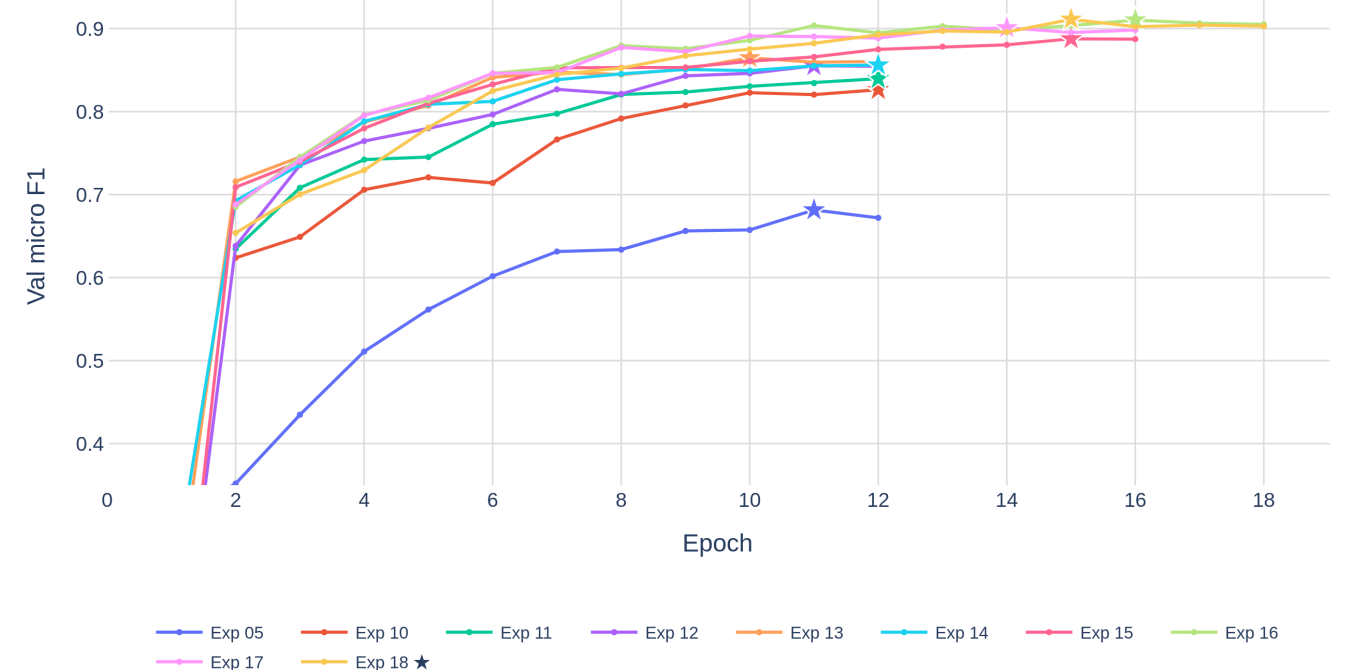


Figure 5. Validation micro F1 across training.

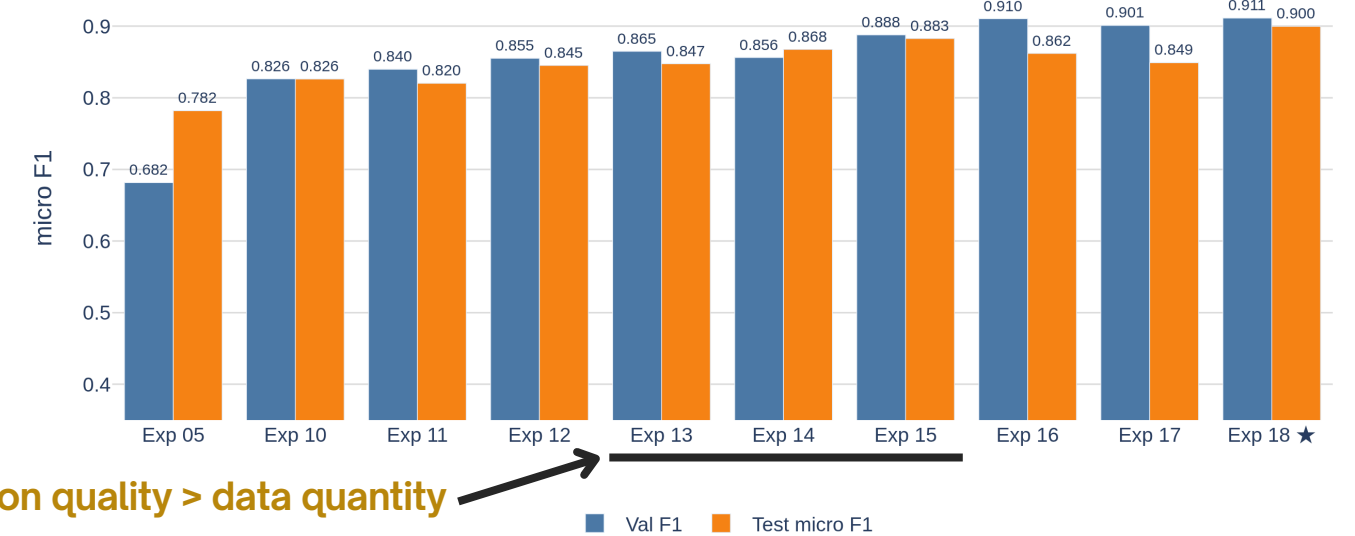


Figure 6. Micro F1 across 18 experiments.

Results

(validation set, per-class threshold)

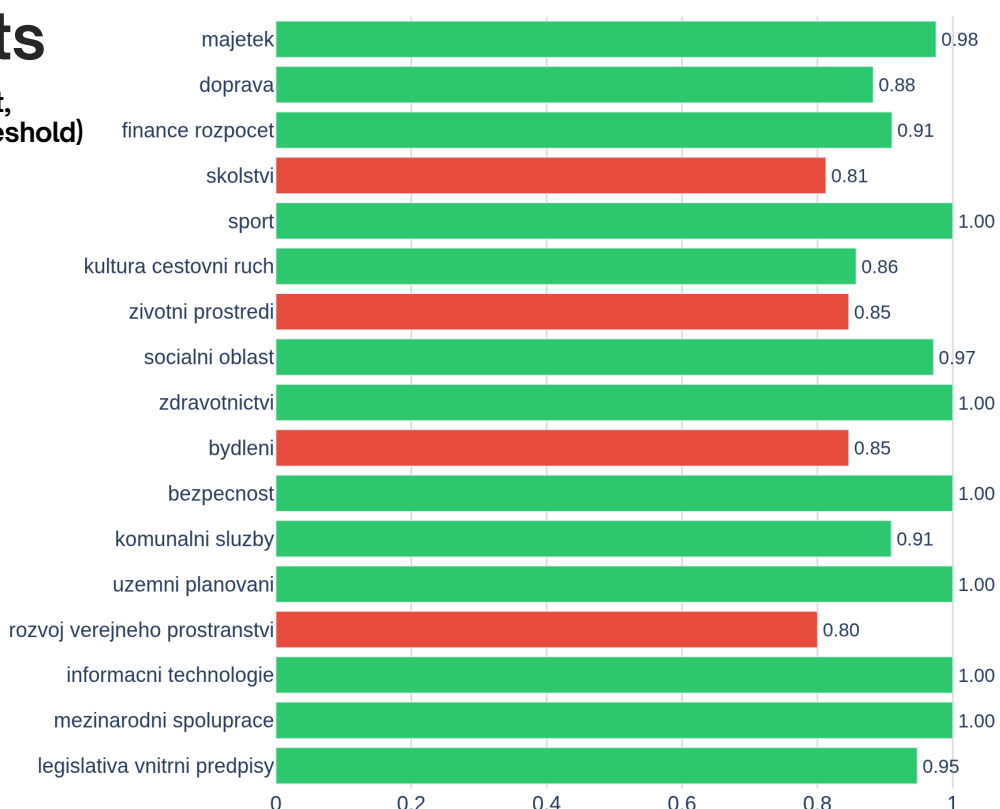


Figure 3. Per-class F1, validation set. Red: F1 < 0.86.



Annotation quality > data quantity

Text materiálu (název + anotace):
Tramvajová trať Plotní – majetkoprávní vypořádání pozemků
Zastupitelstvo souhlasí s nabytím pozemků dotčených stavbou tramvajové trati Plotní do vlastnictví statutárního města Brna. Jedná se o pozemky v k.ú. Trnitá a Komárov.

Používají se per-class thresholdy z validačního setu.

Předikované třídy:

Model is available with an interactive demo:
huggingface.co/spaces/martin0925/council-voting

