



CTU

CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

Automated Fact Checking Based on Czech Wikipedia

Tomáš Mlynář

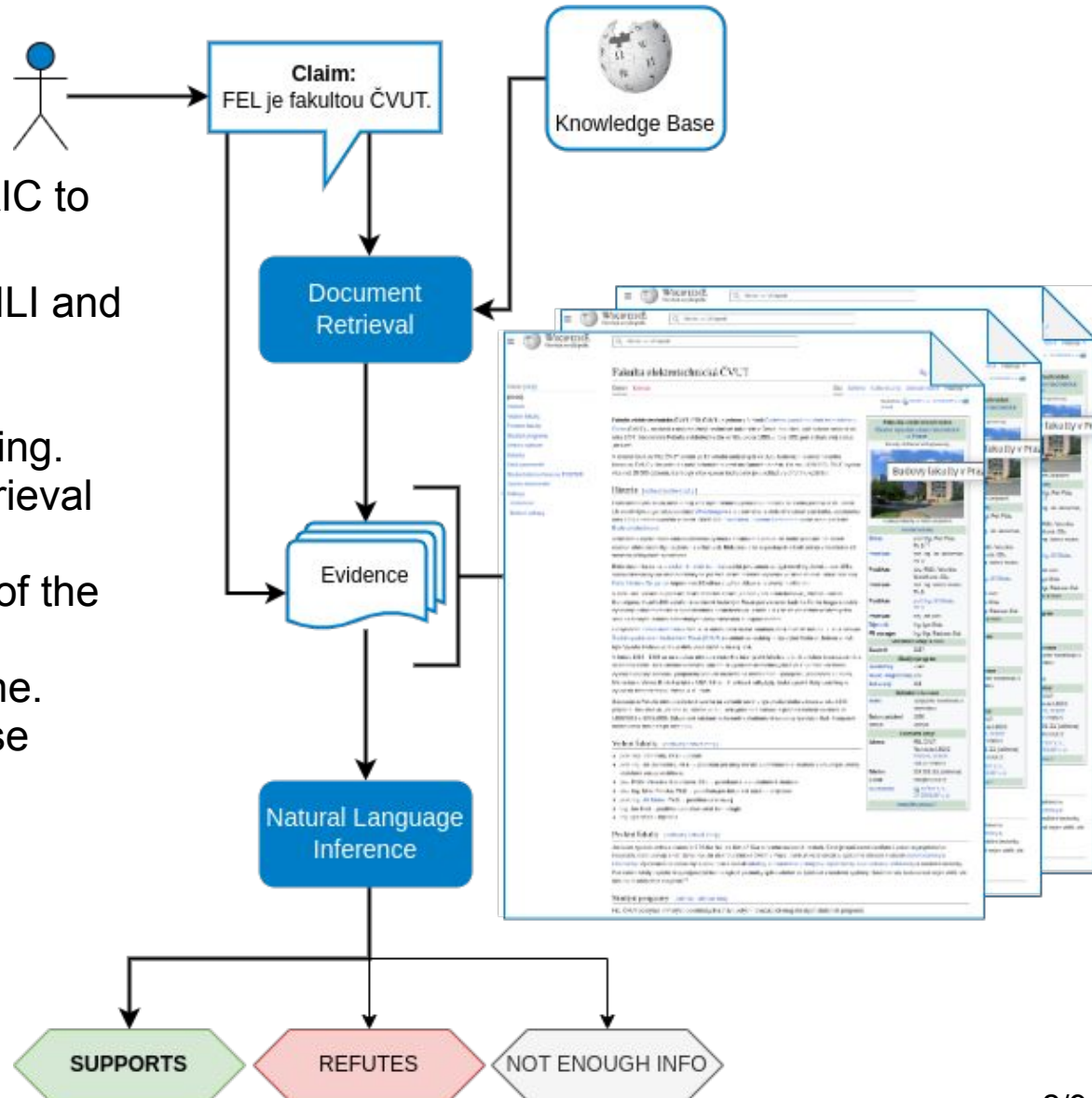
Bachelor's thesis presentation

Supervisor: Ing. Herbert Ullrich

Assignment

Combine previous results from AIC to build a showcase application.

- Explore SOTA methods in NLI and document retrieval.
- Acquire Czech data for Wikipedia-based fact-checking.
- Train NLI and document retrieval models.
- Integrate into initial version of the fact-checking pipeline.
- Evaluate models and pipeline.
- Create a prototype showcase application.



Dataset

1. FEVER dataset
2. Wikipedia dump
3. Translation
4. Mapping of articles
5. Unfold evidence sets

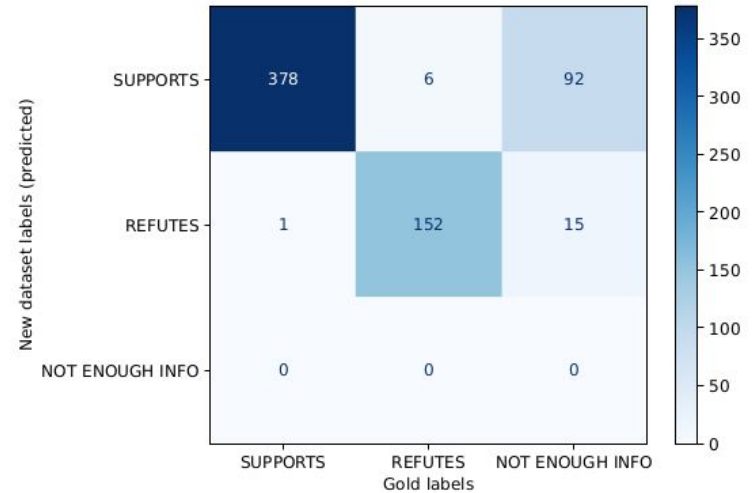
Original FEVER claim	Congressional Space Medal of Honor is the highest award given only to astronauts by NASA.
Facebook	Congressional Space Medal of Honor je nejvyšší ocenění, které NASA uděluje pouze astronautům.
Google	Congressional Space Medal of Honor je nejvyšší ocenění, které NASA uděluje pouze astronautům.
DeepL	Vesmírná medaile cti Kongresu je nejvyšší ocenění, které NASA uděluje pouze astronautům.

```
{  
  "id": 139037,  
  "verifiable": "VERIFIABLE",  
  "label": "SUPPORTS",  
  "claim": "Star Trek: Discovery is a series.",  
  "evidence": [[ [161821, 176242, "Star_Trek-COLON-Discovery", 0] ],  
                [ [161821, 176243, "Star_Trek-COLON-Discovery", 1] ],  
                [ [161821, 176244, "Star_Trek-COLON-Discovery", 6], [16  
↪ 1821, 176244, "Sonequa_Martin-Green", 3] ] ] ]
```

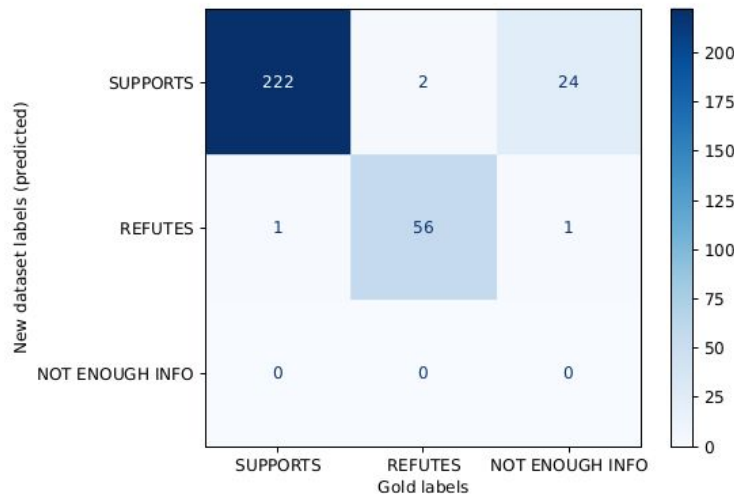
Filtering

1. Predict new labels and scores
2. Temperature scaling
3. Optimize thresholds - F1, precision, 0.7
4. Filter
5. New annotations of 1 % of data points

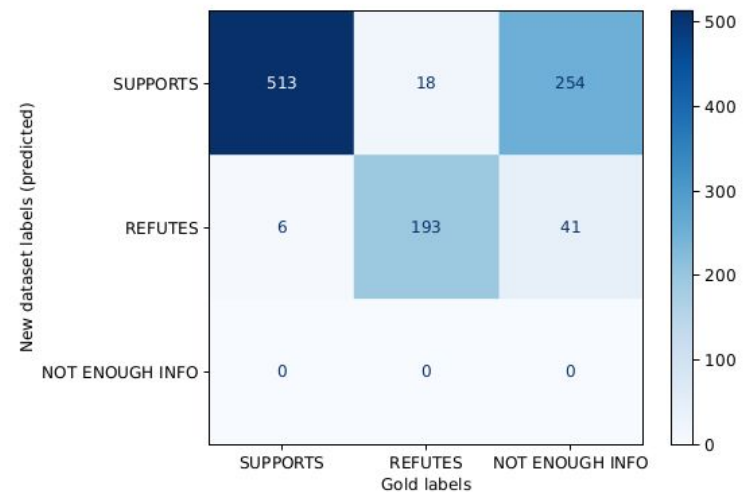
(a): F_1 threshold.



(b): Precision threshold.

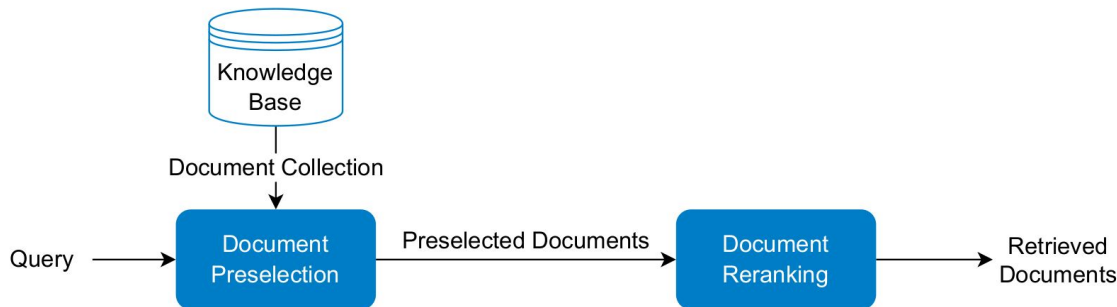


(c): 0.7 threshold.



Document Retrieval

- sparse - Anserini (BM25) - traditional baseline
- hybrid - Anserini+Cross-encoder



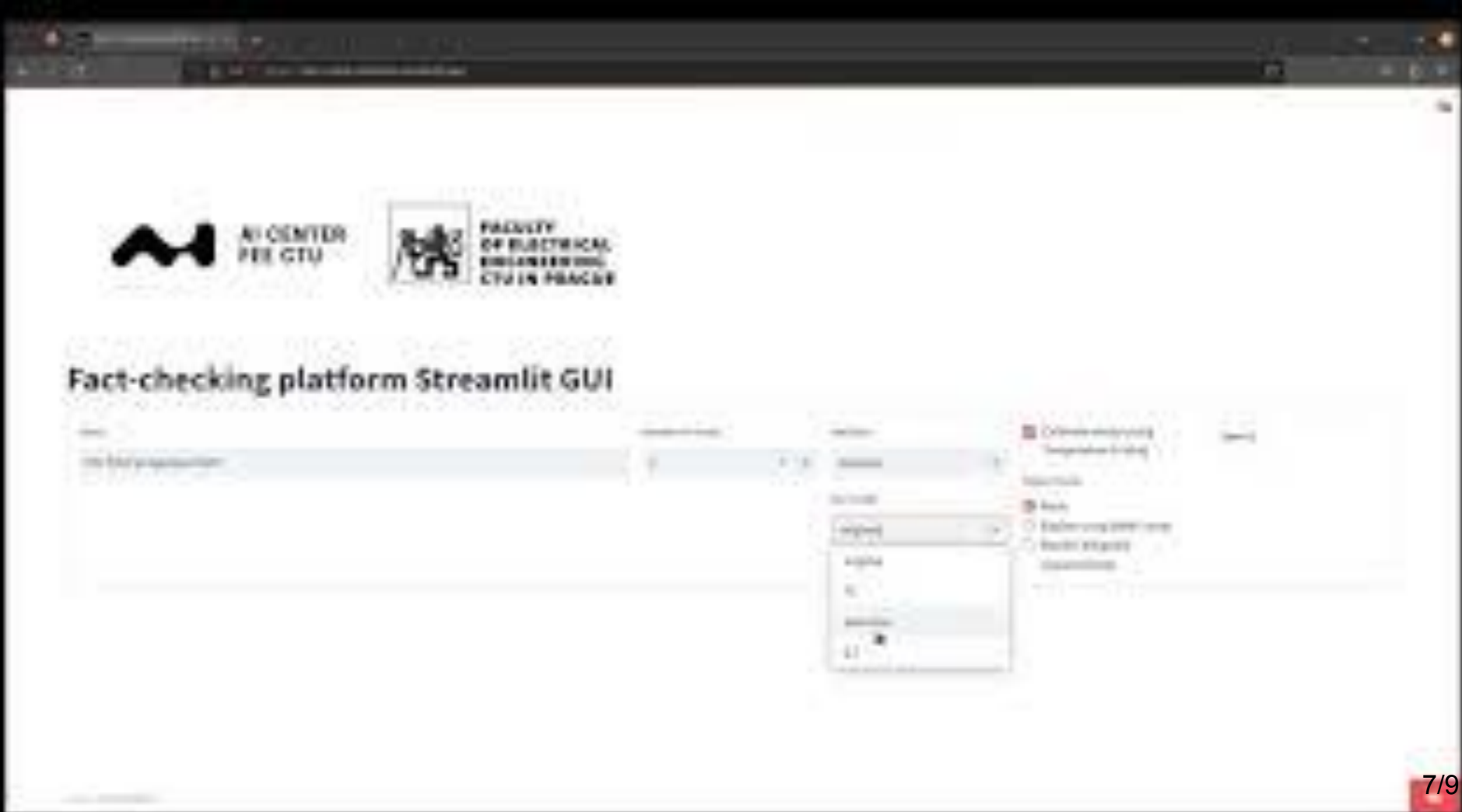
Dataset	Retriever	MRR@1	MRR@5	MRR@10	MRR@20
F_1	Anserini	32.53	43.53	44.87	45.28
	Anserini+Cross-encoder	41.10	50.21	51.29	51.72
Precision	Anserini	30.56	42.16	43.84	44.25
	Anserini+Cross-encoder	41.31	50.63	51.80	52.25
0.7	Anserini	32.92	43.30	44.58	45.02
	Anserini+Cross-encoder	38.51	47.77	48.82	49.26
Noisy	Anserini	29.59	38.5	39.65	40.02
	Anserini+Cross-encoder	37.88	46.50	47.40	47.76

Natural Language Inference

- XLM-RoBERTa Large @SQuAD2 - finetuned by Deepset
- models available on Hugging Face
- F1 macro scores for different models (vertical) and datasets (horizontal):

	Test	CsFEVER-NLI	CTKfactsNLI	ANLI	SNLI
XLM-RoBERTa - Noisy	82.65	56.28	69.70	32.46	36.47
XLM-RoBERTa - F_1	92.49	66.07	58.08	26.26	38.85
XLM-RoBERTa - Precision	94.90	56.00	33.97	18.71	37.80
XLM-RoBERTa - 0.7	89.22	60.00	49.50	27.61	38.09

Prototype Showcase Application



Conclusion

Thesis Accomplishments:

- Explored state-of-the-art methods for NLI and document retrieval.
- Preprocessed Wikipedia dump from Hugging Face.
- Localized FEVER claims using implemented translator and translation model.
- Created new Czech dataset available on Hugging Face.
- Applied noise filtering on three optimized thresholds (with limited success).
- Finetuned and evaluated NLI models on the new datasets (available on Hugging Face).
- Implemented and evaluated Anserini and Hybrid document retrievers.
- Integrated NLI and document retrieval models into a pipeline.
- Developed a prototype showcase application.

Thank you for your attention!

