BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Master's Thesis Assignment

143409

Institut:	Department of Computer Graphics and Multimedia (UPGM)
Student:	**Kamenický Daniel, Bc.**
Programme:	Information Technology and Artificial Intelligence
Specialization:	Software Engineering
Title:	**Designing a Multilingual Fact-Checking Dataset from Existing Question-Answering Data**
Category:	Speech and Natural Language Processing
Academic year:	2022/23

Assignment:

1. Describe the problem of multilingual fact-checking and the importance of using sources from different languages.
2. Research available data sources for multilingual fact-checking and describe their problem formulation.
3. Research how existing question-answering (QA) datasets can be converted into fact-checking (FC) datasets.
4. Design an automatic approach for QA to FC dataset conversion
5. Implement your design.
6. Analyze the properties of the newly converted dataset.
7. Evaluate the difficulty of the problem introduced in your dataset by the baseline model.
8. Create an A1 poster presenting your work.

Literature:

- Thorne, James, et al. "FEVER: a Large-scale Dataset for Fact Extraction and VERification." *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 2018.
- Park, Jungsoo, et al. "FaVIQ: FAct Verification from Information-seeking Questions." *arXiv preprint arXiv:2107.02153* (2021).
- Nørregaard, Jeppe, and Leon Derczynski. "DANFEVER: claim verification dataset for Danish." *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*. 2021.
- SLÁVKA, Michal. Multilingual Open-Domain Question Answering. Brno, 2020. *Master's thesis*. Brno University of Technology, Faculty of Information Technology. Supervisor Ing. Martin Fajčík
- Ruder, S. and Sil, A., 2021, November. Multi-Domain Multilingual Question Answering. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts* (pp. 17-21).

Requirements for the semestral defence:
assignment items 1-4

Detailed formal requirements can be found at https://www.fit.vut.cz/study/theses/

Supervisor:	**Fajčík Martin, Ing.**
Head of Department:	Černocký Jan, prof. Dr. Ing.
Beginning of work:	1.11.2022
Submission deadline:	17.5.2023
Approval date:	2.3.2023