MASTER'S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Ullrich Herbert** Personal ID number: **434653**

Faculty / Institute: **Faculty of Electrical Engineering**

Department / Institute: **Department of Computer Science**

Study program: **Open Informatics**

Specialisation: **Artificial Intelligence**

II. Master's thesis details

Master's thesis title in English:

Dataset for Automated Fact Checking in Czech Language

Master's thesis title in Czech:

Datová sada pro automatizované ověřování faktů v českém jazyce

Guidelines:

The task is to develop a methodology for collecting a Czech dataset aimed at fact-checking. Collect and process the data and finally perform initial experiments with textual entailment recognition.
1) Explore state-of-the-art in automated fact-checking. Focus on 1) dataset collection methodologies, 2) the last stage of the fact-checking pipelines, i.e., methods of textual entailment recognition.
2) Develop a methodology for Czech fact-checking dataset collection based on the related FEVER [1] Wikipedia-based dataset. The source of data for the new dataset will be the Czech News Agency (ČTK).
3) Create or modify an existing tool and use it to collect dataset.
4) Perform exploratory data analysis.
5) Use the dataset to develop and experiment with initial textual entailment recognition models. These models aim to classify whether textual claims are supported or refuted w.r.t. other reference documents. Evaluate the models using standard approaches used in textual entailment recognition.

Bibliography / sources:

[1] Thorne, James, et al.:FEVER: a large-scale dataset for fact extraction and verification; arXiv preprint arXiv:1803.05355 (2018).
[2] Thorne, James, et al.:The fact extraction and verification (fever) shared task; arXiv preprint arXiv:1811.10971 (2018).
[3] Binau, Julie, and Henri Schulte: Danish Fact Verification: An End-to-End Machine Learning System for Automatic Fact-Checking of Danish Textual Claims; (2020).
[4] Poliak, Adam: A survey on recognizing textual entailment as an NLP evaluation." arXiv preprint arXiv:2010.03061 (2020).
[5] Storks, Shane, Qiaozi Gao, and Joyce Y. Chai: Recent advances in natural language inference: A survey of benchmarks, resources, and approaches; arXiv preprint arXiv:1904.01172 (2019).

Name and workplace of master's thesis supervisor:

Ing. Jan Drchal, Ph.D., Department of Theoretical Computer Science, FIT

Name and workplace of second master's thesis supervisor or consultant:

Date of master's thesis assignment: **09.02.2021** Deadline for master's thesis submission: _____

Assignment valid until: **30.09.2022**

Ing. Jan Drchal, Ph.D. Supervisor's signature	Head of department's signature	prof. Mgr. Petr Páta, Ph.D. Dean's signature

III. Assignment receipt

The student acknowledges that the master's thesis is an individual work. The student must produce his thesis without the assistance of others, with the exception of provided consultations. Within the master's thesis, the author must state the names of consultants and include a list of references.

.	
Date of assignment receipt	Student's signature