MASTER‘S THESIS ASSIGNMENT

I. Personal and study details

Student's name: **Dvořáček Dominik** Personal ID number: **492169**

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:

Multi-document Structured Summarization

Master’s thesis title in Czech:

Metody strukturované sumarizace množiny dokumentů

Guidelines:

This thesis focuses on exploring methods for multi-document structured summarization. The emphasis will be on summarization techniques rather than document retrieval. The structured component of the summarization will involve knowledge-extraction tasks, such as determining the temporal relevance of events and extracting information about entities like people, organizations, and places.
1) Conduct a review of existing large language model (LLM)-based methods for multi-document summarization. Additionally, explore prompt-based approaches for knowledge extraction.
2) Analyze a news corpus dataset provided by the supervisor.
3) Develop an NLP pipeline to summarize a selected subset of the corpus. The document sets will be chosen based on tags related to a topic of interest, so document retrieval is not required.
4) Design an experimental methodology, run experiments, and evaluate the summarization results.

Bibliography / sources:

[1] Li, Qian, et al. "A survey on deep learning event extraction: Approaches and applications." IEEE Transactions on Neural Networks and Learning Systems (2022).
[2] Ma, Congbo, et al. "Multi-document summarization via deep learning techniques: A survey." ACM Computing Surveys 55.5 (2022): 1-37.
[3] Liu, Pengfei, et al. "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing." ACM Computing Surveys 55.9 (2023): 1-35.
[4] Nasar, Zara, Syed Waqar Jaffry, and Muhammad Kamran Malik. "Named entity recognition and relation extraction: State-of-the-art." ACM Computing Surveys (CSUR) 54.1 (2021): 1-39.
[5] Steen, Julius, and Katja Markert. "Abstractive timeline summarization." Proceedings of the 2nd Workshop on New Frontiers in Summarization. 2019.

