



## ZADANIE ZÁVEREČNEJ PRÁCE

- Meno a priezvisko študenta:** Bc. Viktória Ondrejová  
**Študijný program:** aplikovaná informatika (Jednoodborové štúdium, magisterský II. st., denná forma)  
**Študijný odbor:** informatika  
**Typ záverečnej práce:** diplomová  
**Jazyk záverečnej práce:** anglický  
**Sekundárny jazyk:** slovenský
- Názov:** Exploring Translation and Reasoning Abilities of Large Language Models in the Saris Dialect  
*Skúmanie prekladateľských a uvažovacích schopností veľkých jazykových modelov v šarišskom dialekte*
- Anotácia:** Táto práca bude skúmať možnosti prekladu a uvažovania veľkých jazykových modelov (LLM) v rámci šarišského dialektu. Na vývoj prekladových modelov medzi Šarišom a slovenčinou bude využívať LLM v konfiguráciách zero-shot a data augmentation. Rodení hovoriaci overia presnosť modelu pomocou manuálneho hodnotenia. Pre úlohy uvažovania bude pripravený nový súbor údajov, ktorý spolu s existujúcim pomôže posúdiť schopnosti kauzálneho uvažovania LLM v dialekte.
- Cieľ:** Ciele diplomovej práce zahŕňajú (ale nie sú obmedzené na)
- Analýza existujúcich najmodernejších metodológií strojového prekladu a uvažovania zdravým rozumom so zameraním na jazyky a dialekty s nízkymi zdrojmi.  
Vývoj prekladateľských modelov uľahčujúcich presnú konverziu medzi šarišským dialektom a spisovnou slovenčinou pomocou LLM v konfiguráciách zero-shot a data augmentation.  
Spracovanie a/alebo vytváranie súborov údajov zameraných na kauzálne zdôvodnenie a prekladateľské úlohy pre šarišský dialekt, pričom sa zabezpečí lingvistická a kontextová relevantnosť.  
Návrh a implementácia prístupov strojového prekladu a uvažovania prispôbených jazykovým nuansám šarišského dialektu s využitím schopností LLM.  
Analýza a interpretácia výkonnosti vyvinutých modelov, identifikácia kľúčových poznatkov a vzorcov chýb na zlepšenie spracovania prirodzeného jazyka špecifického pre dialekty.
- Literatúra:** Min, Bonan, et al. "Recent advances in natural language processing via large pre-trained language models: A survey." ACM Computing Surveys 56.2 (2023): 1-40.  
Ebing, Benedikt, and Goran Glavaš. "To Translate or Not to Translate: A Systematic Investigation of Translation-Based Cross-Lingual Transfer to Low-Resource Languages." arXiv preprint arXiv:2311.09404 (2023).



Univerzita Komenského v Bratislave  
Fakulta matematiky, fyziky a informatiky

---

Tars, Maali, Andre Tättar, and Mark Fišel. "Extremely low-resource machine translation for closely related languages." arXiv preprint arXiv:2105.13065 (2021).

**Vedúci:** Mgr. Marek Šuppa  
**Katedra:** FMFI.KAI - Katedra aplikovanej informatiky  
**Vedúci katedry:** doc. RNDr. Tatiana Jajcayová, PhD.

**Spôsob sprístupnenia elektronickej verzie práce:**  
bez obmedzenia

**Dátum zadania:** 07.05.2024

**Dátum schválenia:** 13.05.2024

prof. RNDr. Roman Ďurikovič, PhD.  
garant študijného programu

---

študent

---

vedúci práce