

Slovenský národní korpus: exkurze, Bratislava

Datum konání: 25. 11. 2022

Místo konání: Slovenský národní korpus, Jazykovědný ústav L. Štúra SAV, v. v. i., Panská 26, 811 01 Bratislava 1

Název přednášky: Slovenský národní korpus: exkurze, Bratislava

Přednášející: Mgr. Ivor Uhliarik

Počet účastníků: 20

Zpracovali: Dědič Martin, Sasínek Tomáš

Biogram

Mgr. Ivor Uhliarik je vedoucím IT složky Slovenského národního korpusu. Vystudoval informatiku a umělou inteligenci na Univerzitě Komenského v Bratislavě. Momentálně na téže univerzitě dodělává doktorské studium. Součástí SNK se stal v roce 2015, kde začínal jako NLP výzkumník. Předtím také pracoval jako vývojář a AI výzkumník.

Závěr

Přednáška jako taková se odehrávala v nejvyšším patře budovy Slovenského národního korpusu, které vypadalo moderněji než ostatní patra budovy. Jelikož přednášející je především informatik, přednáška se dívala na lingvistiku z tohoto pohledu. Mluvílo se o činnostech SNK, o korpusech a jejich druzích, databázích a nástrojích, které jsou používány v SNK.

Velice nás zaujala publikace *Frekvenčního slovníku hovorovej slovenčiny*, ve kterém jsme mohli vidět, jak často se jednotlivá slova ve slovenštině používají. Také jsme byli rádi za osvěžení paměti o word embeddings, ve kterých můžete hledat jednotlivá slova pomocí vztahu s jiným slovem. Nástroj, který si získal naši pozornost, je slovenský SpaCy, který je naprogramován v Cythonu, což je jazyk založený na Pythonu s prvky z jazyku C, díky čemuž využívá výhod nízkourovňového programovacího jazyka. Velmi zajímavý byl závěr přednášky, kde nám bylo řečeno, co potřebují počítačové lingvisté za znalosti, aby mohli pracovat ve Slovenském národním korpusu. Kromě znalostí jazyka Python je také potřebná například znalost unixového prostředí.



