

S rostoucím objemem dat, zejména nestrukturovaného textu, roste důležitost zpracování přirozeného jazyka. Nejmodernějšími technologiemi posledních let jsou neuronové sítě. Tato práce aplikuje nejúspěšnější metody, jmenovitě Bidirectional Encoders Representations from Transformers (BERT), na tři české úlohy ve zpracování přirozeného jazyka – lematizaci, morfologické značkování a analýzu sentimentu. Použili jsme BERTa s jednoduchou klasifikační hlavou na tři české dataset pro analýzu sentimentu: mall, facebook a csfd a dosáhli jsme state-of-the-art výsledků. Také jsme prozkoumaly několik možných postupů trénování pro úlohy značkování a lematizace a obdrželi jsme nové state-of-the-art výsledky pro Pražský závislostní korpus v obou úlohách pomocí fine-tunningu. Konkrétně jsme dosáhli přesnosti 98.57% pro značkování, 99.00% pro lematizaci a 98.19% pro společné ohodnocení. Nejlepší modely pro všechny úlohy jsou veřejně dostupné.