

Abstrakt: Rodina úkolů pro zpracování přirozeného jazyka (NLP), jako je označování po částech řeči (PoS), identifikace pojmenované entity (NER) a identifikace více slov (MWE), zahrnují přiřazení štítků sekvencím slov v textu označování). Většina moderních přístupů strojového učení k sekvencímu označování využívá vkládání slov, naučené reprezentace textu, ve kterých mají slova s podobnými významy podobné reprezentace. Docela nedávno, kontextualizované slovní embeddings získaly hodně pozornosti, protože na rozdíl od předem vyškolených kontextově necitlivých embeddings, jako je word2vec, jsou schopni zachytit význam slova v kontextu. V této diplomové práci hodnotím výkonnost různých nastavení vkládání (kontextu citlivé, kontextově necitlivé slovo, stejně jako slovo specifické pro danou práci, charakter, lemma a PoS) na třech výše uvedených úlohách označování sekvence pomocí hlubokého modelu učení ( BiLSTM) a portugalských datových sad.