

Tato bakalářská práce komplexně zkoumá hodnocení důležitosti slov, od definování tohoto pojmu po vytvoření a vyhodnocení predikčního systému. Pomocí našeho webového anotačního nástroje jsme sebrali ruční odhady důležitosti slov; důležitost přitom definujeme jako relativní uspořádání slov. Navrhujeme metodu self-supervised strojového učení, kde jsou do textu uměle vložena nová slova a my pak doladujeme model BERT, aby se naučil tato slova identifikovat. Předpokládáme, že výsledný model přidělí vyšší pravděpodobnost vložení méně důležitým slovům. Experimentujeme se dvěma různými strategiemi vkládání: metodou vkládání seznamem a metodou vkládání BERTem. Vyhodnocení na našich shromážděných datech ukazuje, že naše metody překonávají tradiční základní metody jako TF-IDF a soupeří s existujícími přístupy, což dokládá funkčnost našeho přístupu při predikci důležitosti slov.