

Práce se zabývá porovnáváním textu a jeho kategorizaci. Kategorie, které je program schopen určit, získává v módu učení. Porovnává několik možných algoritmů, které lze využít ke kategorizaci textu. Jde především o Bayesovský model, klasifikaci pomocí neuronových sítí a vektorový model. V praktické části je implementován vektorový model, který využívá kosinovu míru podobnosti. Extrakce termínu vychází z Luhnovy myšlenky o významovosti slov. Jako hlavní zdroj vah pro kosinovu míru podobnosti je využívána hlavně metoda tfidf s penalizacemi.