

V kontextu analýzy diskurzu stupeň aktivovanosti (salience) modeluje aktuální míru zapojenosti odkazovaných objektů a její vývoj v průběhu textu. Algoritmus pro určování aktivovanosti a vizualizaci jejího průběhu již byl navržen a otestován na malém vzorku dat. Tato práce reprodukuje výsledky algoritmu ve větším měřítku pomocí dat z Pražského závislostního korpusu 3.0. Výsledky jsou pak zpracovány do přístupného tvaru a je provedena jejich analýza jak pomocí vizuálního výstupu, tak i výstupů kvantitativních. Přitom jsou zohledněny dva základní stavební kameny aktivovanosti; koreferenční vztahy a informační struktura věty. Práce se také zabývá možností modelování aktivovanosti pomocí strojového učení za použití algoritmů rozhodovacích stromů a náhodného lesa. V závěrečné části je proveden experiment zkoumající možné využití informace o aktivovanosti v některé z úloh strojového učení při zpracování přirozeného jazyka na příkladě shlukování dokumentů.