

Abstrakt

Vedle dotazníkových šetření představují texty důležitý zdroj dat pro sociologický výzkum již od počátků jeho rozvoje. Metody analýzy textů v sociologii zahrnují dva základní vývojové proudy: první představuje kvantitativní obsahová analýza Bernarda Berelsona, druhý hermeneutická analýza Hans-Georga Gadamera. V posledních dvaceti letech jsou pak oba tyto metodologické proudy ovlivněny rozvojem informačních technologií. Předkládaná práce se zabývá jednou z metod počítačové textové analýzy (CATA), která stojí na pomezí obou těchto metodologických proudů, metodu sledování spoluvýskytu slov. Práce představuje tuto metodu v kontextu ostatních metod analýzy textu a zmiňuje se i o inspiračních zdrojích dalšího rozvoje těchto metod, o korpusové lingvistice a text miningu. Ve druhé části pak rozebírá jednotlivé kroky analýzy spoluvýskytů slov v textu: sestavení textového korpusu, sestavení slovníku, výpočet datové matice a vizualizace vzdáleností slov s užitím metody mnohorozměrného škálování. Metoda je dále aplikována na konkrétní data, dva textové korpusy sestavené z přepisů biografických interview s aktéry československé normalizace, s disidenty a komunistickými funkcionáři. U těchto korpusů je posouzena kvalita modelů v závislosti na volbě parametrů (koeficient vzdálenosti, velikost kontextové jednotky). Následně jsou tyto modely interpretovány. Pro posouzení validity jsou tyto interpretace dále konfrontovány s výsledky kvalitativní hermeneutické analýzy provedené na stejných datech.