

<b>Tytuł projektu:</b>	<b>Odkrywanie wiedzy w dużych zbiorach publikacji naukowych z zakresu socjologii</b>
<b>Skrócony opis:</b>	<p><b>Celem</b> projektu jest odkrywanie tendencji w tematach publikowanych w ramach socjologii. Na potrzeby projektu zbudowana została metoda pozwalająca na badania i eksploracje dużych zbiorów danych tekstowych. W pierwszym etapie projekt realizowany jest dla analiz w języku angielskim, a w kolejnym dla języka polskiego.</p> <p><b>Realizacja celów</b> badawczych projektu opiera się na: zbudowaniu procedur dla wielowymiarowych analiz text minigowych; skonstruowaniu korpusu tekstów służących do identyfikacji słownictwa i języka typowego dla publikacji z zakresu socjologii. Analizie poddawane są: encyklopedie i słowniki socjologii jak i publikacje z czasopism zakwalifikowanych do listy filadelfijskiej. Projekt realizowany jest poprzez wykorzystanie: technik specyficznych dla analiz jakościowych, technik <i>Natural Language Processing</i>, algorytmów uczenia maszynowego oraz klasyfikacji treści za pomocą podejścia <i>Knowledge-Based Systems</i> i <i>Dictionary-Based Classification Models</i>.</p> <p>Projekt ma kilka <b>praktycznych aplikacji</b> dla badań i analiz tekstów. Poprzez odkrywanie wiedzy w tekście możliwe jest kwalifikowanie publikacji do danego obszaru tematycznego (lub określenie subdyscypliny). Rozpoznanie tematyki z kolei pozwala na określenie trendów w publikacjach z danego obszaru, w danym okresie czasu.</p>
<b>Kierownik projektu:</b>	Krzysztof Tomanek (współpraca Paweł Mielniczek, Marcin Tuleja)