

V této práci se snažím nalézt definici zpravodajského tématu tak, aby byla detekce těchto témat v textu implementovatelná a kvalita této detekce měřitelná. Popisuji možné metody — „prosté“ počítání slov, případně se zavedením stopslov; TF-IDF; dále popisuji problém textové klasifikace, mírně se dotknu text clusteringu. Dále popisuji přístupy, nazvané latent semantic indexing a latent Dirichlet allocation. Také popisuji experimenty s „prostým“ počítáním slov, TF-IDF a textovou klasifikací na databázi článků z několika elektronických zdrojů; vznik této databáze v práci popisuji rovněž. Ke způsobu řešení pomocí textové klasifikace uvádím metriku pomocí měření přesnosti a úplnosti; podle těchto metrik měřím několik variant textové klasifikace.