

Možnost automaticky sledovat popularitu osob v novinách by jistě uvítaly nejen tyto osoby samotné. Počítačové zpracování subjektivity je sice rychle se rozvíjející podobor počítačové lingvistiky, v češtině ovšem vůbec pro analýzu subjektivity a polaritu v publicistice neexistují data. Začali jsme tedy s tvorbou ručně anotovaného korpusu polarit českých publicistických textů, které se ovšem pro takové zpracování ukázaly jako krajně nevhodné. Dále jsme navrhli klasifikátor založený na statistických metodách, který by měl na základě tohoto korpusu popularitu sledovat, a otestovali jsme ho na korpusu recenzí bílého zboží a orientačně na záznamu našeho korpusu vět z novinových článků. Jako model jsme použili automaticky extrahovaný unigramový slovník, tři příbuzné metody pro zjišťování polárních lemmat a množství filtrů pro selekci relevantních lemmat. Na recenzích bílého zboží jsme dosáhli výsledků srovnatelných se světovým výzkumem už se základním modelem, naopak u českých publicistických textů vidíme kvůli jejich charakteru možný příslib až u více lingvisticky orientovaných metod.