

Seminární práce předmětu CJBB85 Počítačové nástroje pro češtinu

Seminární práce souvisí s potřebami forenzní lingvistiky. Téma bylo připraveno ve spolupráci s Mgr. Veronikou Novákovou z Kriminálního ústavu v Praze.

Úkolem je pokusně identifikovat slova, která se mohou potenciálně vyskytovat ve výhružných dopisech (e-mailech, sms) a propagandistických, manipulativních, rasistických či jinak nebezpečných textech na internetu.

- 1) najděte vhodný zdroj informací – webové stránky, diskuze, sociální sítě – Facebook, Twitter apod., pro výběr sloves je možné také využít sémantické třídy sloves na <https://nlp.fi.muni.cz/~ydana/Semtridy.html>
- 2) najděte minimálně 5 substantiv, 5 adjektiv a 5 sloves, která se ve vybraných textech často objevují (maximální počet slov není stanoven)
- 3) v korpusech SYN v5 a/nebo CzTenTen12 zjistěte frekvenci vybraných slov, hledejte přes lemma všechny tvary slov
- 4) najděte několik nejvýstižnějších a nejfrekventovanějších slovních spojení ve Word Sketch, která tematicky spadají do zkoumané oblasti
- 5) do textového dokumentu (např. ve Wordu) sepište:

zdroje textů, nalezená slova (uspořádaná podle slovních druhů) s uvedenou frekvencí (a názvem korpusu, kde jste frekvenci zjišťovali), nejzajímavější slovní spojení
- 6) práci odevzdejte do odevzdáárny předmětu CJBB85 v IS do **20. 1. 2018**