

Princíp fungovania

Matúš Mladý

Skript prečíta zadaný text pričom každé slovo dá do tvaru lemy (tvar v ktorom sa nachádza v slovníku, podobné ako koreň slova stem). Následne spočíta koľko daných lem sa nachádza v celom texte. Toto počítanie priradí každému slovu váhu (výskyt podstatného mena 10, cudzieho slova 9, slovesa 8, prídavného mena 7), ostatné slová ignorujeme, dajú sa považovať za stop-words teda slová bez významu. Slovám prirátame 1 bod váhy za každé 3 písmená v slove (dlhšie slová môžu mať väčší význam). Takto má po analýze textu každé jedinečné slovo (vynímajúc bezvýznamové slová) nejakú váhu podľa toho aký je slovný druh, aké je dlhé ale predovšetkým podľa toho ako často sa v texte vyskytuje. Ako výsledok vypíšeme všetky slová s váhou aspoň 30% váhy najdôležitejšieho slova. Takto zabezpečíme relevantnosť slov či už v dlhších alebo aj kratších textoch.