

Když se řekne klíčové slovo, tak mě nejdřív napadne těch pár klíčových slov, která jsou uvedena u každého článku nebo práce a která obvykle autor vybírá tak, aby co nejlépe vystihovala oblast, které se daný text týká. Jistě bychom si dokázali představit i nějaký algoritmus, který by taková slova z článku vybíral a jehož výstupem by bylo právě těch 5 klíčových slov nebo slovních sousloví.

Otázka je, jestli by tato slova vybral program lépe než autor textu a jestli vůbec lze správně zařadit nějaký text pomocí pěti slov. Protože když jsem se zkoušel dívat v ISu na bakalářské práce s podobným zaměřením, jako má ta moje (tato podobnost se určuje právě podle klíčových slov), tak mi systém vracel spíše úplné nesmysly a práce s tou mou nesouvisející. Nejjednodušší vysvětlení je to, že jsme buď já, nebo autoři oněch dalších prací vybrali klíčová slova špatně. Ovšem jiný důvod by mohl být ten, že výběrem pouhých několika slov se ztrácí velká informace o textu. Tato informace by mohla být uchovávána a využívána lépe – programy by si mohly místo několika slov pamatovat například frekvence výskytu všech slov, každé slovo charakterizovat dalšími parametry nebo si pamatovat dokonce vazby mezi jednotlivými slovy v textu. Takováto struktura by lépe charakterizovala daný článek. S tímto problémem by souviselo také to, jak takovéto výsledky mezi sebou porovnávat. Každý autor může psát jiným stylem, používat jiná slova nebo jim přiřkládat odlišný význam. S rozvojem internetu a digitalizace dat se objevují nové možnosti využití klíčových slov. Je rozdíl mezi novinovým článkem a například webovou stránkou, kde existují nadpisy, odkazy a podobně. Tyto informace by měly algoritmy pracující s klíčovými slovy využívat a slovům v jiných částech stránky přiřkládat například jinou váhu. Zároveň lze využívat i informací, kam vedou odkazy z dané stránky a odkud naopak na ni. Zde vidím možnosti uplatnění algoritmů pracujících s klíčovými slovy.