

Posudek oponenta diplomové práce

Autor/autorka práce: **Václav Rajtmajer**

Název práce: **Sledování trendů na sociální síti Twitter**

Diplomová práce se zabývá zpracováním příspěvků ze sociální sítě Twitter. Cílem je detekovat významné události, na které uživatelé Twitteru reagují. To je provedeno pomocí shlukování souvisejících příspěvků. Na základě analýzy shluků jsou pak vybrány důležité události. Výsledný systém umožňuje detekci událostí v reálném čase.

Teoretická část je věnována popisu sociální sítě Twitter a nástrojů umožňujících získat data a strojově je zpracovat. Dále jsou zde popsány základní techniky zpracování a klasifikace textu. Praktická část pak popisuje experimentální systém, který byl v průběhu práce implementován a experimentální ověření výsledků na korpusu, který byl pro tyto účely vytvořen a ručně anotován.

Prezentovaný systém přináší zajímavé výsledky. Je zde obtížné objektivně zhodnotit dosažené výsledky, protože se jedná o relativně novou úlohu a nejsou k dispozici další studie na toto téma v českém jazyce. Nicméně vzhledem k charakteru dat z Twitteru se 50 % přesnost jeví jako slibný výsledek.

Práce je napsána solidní formou, ale některé věci by mohly být formulovány jasněji. Bylo by také vhodné důkladněji zkontrolovat pravopisné chyby a překlepy, kterých je zde poměrně hodně. Ocenil bych také číslování rovnic. Některé části jsou nevhodně uspořádány. Například popis vyhodnocení výsledků je až v kapitole 6, ale v textu kapitoly 5 je už vyhodnocení použito.

Algoritmus 1 na straně 21 by možná nebylo nutné uvádět, protože je jednoduchý, oproti tomu algoritmus vytvoření shluku by si zasloužil detailnější popis a zápis v pseudokódu by porozumění ulehčil. Na straně 20 je nekonzistentní označení dokumentu a tweetu.

Přestože je práce spíše výzkumného charakteru, očekával bych detailnější popis implementovaného systému. Rovněž mi zde chybí alespoň stručná uživatelská příručka s instrukcemi, jak přiložený program spustit.

Práci s literaturou hodnotím jako nadstandardní, protože vzhledem k charakteru a novosti úlohy bylo nutné nastudovat značné množství materiálů.

Zadání bylo splněno bez výhrad.

Dotazy k práci

Sekce 3.5.4 - Jaký je počet důležitých slov použitých pro reprezentaci události?

Jaký je rozdíl v interpretaci kosínové vzdálenosti oproti dalším dvěma použitým vzdálenostem?

Navrhuji hodnocení známkou **velmi dobře** a práci doporučuji k obhajobě.

V Plzni 2.6.2015

Ing. Ladislav Lenc, Ph.D.