

Posudek diplomové práce „Advanced Methods for Sentence Semantic Similarity. Ptáček, T. (2012)“

Ing. Ivan Habernal, vecoucí práce
KIV, FAV, ZČU
habernal@kiv.zcu.cz

4. června 2012

Bc. Tomáš Ptáček se ve své práci zabývá metodami porovnávání vět na základě jejich sémantiky. Práce je psaná anglicky, má rozsah 44 stran textu a DVD s programovou součástí. Téma práce úzce souvisí s výzkumným projektem skupiny LIKS a americké firmy OwenSoftware a její výsledky budou v pilotních částech projektu použity.

V teoretické části diplomant dává svoji práci do kontextu Information Retrieval, stejně tak popisuje různé metody pro sémantickou podobnost slov a vět. Taktéž představí dvě veřejné aktuální anglické datové sady pro testování sémantické podobnosti vět (MSRP a STS) a metodiky měření úspěšnosti.

V praktické části je pak implementováno několik state-of-the-art metod popsaných v předchozí části a jejich výsledky jsou ověřeny jak podle původních zdrojů (konferenční články a články v časopisech) tak i na MSRP a STS datech. Inspirován těmito metodami pak diplomant navrhuje metody vlastní ve snaze dosáhnout lepších výsledků na daných datech.

Všechny nově navržené metody jsou otestovány na dodaných datech, přičemž argumentace výsledků je vždy důsledná a odhaluje vlastnosti daných metod. To může pomoci odhalit slabiny metod např. při testování na principiálně jiných datech.

Dva hlavní přínosy práce jsou následující. Zaprvé, vyvinutým metodám se daří překonávat vybrané state-of-the-art metody. Zadruhé, data STS byla během práce vyhodnocena a její výsledky publikovány (jednalo se o tzv. shared task). Ačkoliv se diplomant soutěže neúčastnil, poměřil své metody s výsledky soutěže. V polovině případů se mu podařilo velmi významně překonat baseline úlohy (o 25–50% rel.) a v jednom případě byl pak jeho výsledek téměř srovnatelný s výhercem tasku.

Shrnutí. Diplomant vytvořil významnou práci vědeckého charakteru, na úloze pracoval samostatně, sestudoval existující konferenční a časopisecké články k tématu a přicházel s originálními nápady. Prozkoumal oblast srovnání vět na základě jejich sémantiky a navrhl několik nových metod, jejichž úspěšnost ověřil na veřejně dostupných datech. Proto **doporučuji práci k obhajobě** a hodnotím známkou **v ý b o r n ě**.



Ing. Ivan Habernal

Dodatečné otázky

1. Je možné zjistit, v čem se (v principu) liší metoda výherce STS tasku a diplomantova přístup? (*Pozn.: Pokud výherce STS tasku ještě nepublikoval své výsledky, je možno tuto otázku ignorovat.*)