

Posudek vedoucího diplomové práce

---

Lukáš Hain: Experimenty s Apache Mahout

---

Diplomantovým úkolem bylo ověřit, do jaké míry lze použít nástroj Apache Mahout pro řešení úloh, kterými se zabývá laboratoř zpracování přirozeného jazyka na KIV. Mahout je navržen pro distribuované výpočty a velká data. Tyto pozitivní vlastnosti s sebou přináší i značná negativa. S Mahoutem je obtížné pracovat, je například potřebné manipulovat se speciálním distribuovaným souborovým systémem.

Diplomant v rámci své práce úspěšně vyřešil tři netriviální úlohy z oblasti zpracování přirozeného jazyka. Ve své práci důkladně popisuje svá řešení tak, aby bylo možné jeho experimenty replikovat nebo použít na příbuznou oblast. Diplomátovy výsledky mají vysoký potenciál, že budou využity ve výzkumu. Jedná se zejména dobrý výkon při shlukování. Výsledky pro pracoviště CIV vypadají také slibně. Vzhledem k tomu, že šlo o ostrá data, je zde malé riziko, že by je CIV nemohl využít.

Pro ověření návodu v práci jsem se experimentálně pokusil replikovat poslední úlohu - shlukování vektorů sémantického prostoru na jiných datech. Ověřil jsem, že návod je přesný a podrobný a výpočet shlukování se mi podařilo spustit. Výpočet se bohužel zastavil na 8-mé iteraci. To však není chyba diplomanta, ale nástroje Mahout.

Práce je napsaná v sázecím programu L<sup>A</sup>T<sub>E</sub>X. Je zřejmé, že diplomant na tento program není zvyklý a v práci se dopustil několika formální prohřešků. Například neodkazuje tabulky, na straně 65 je neexistující odkaz na kapitolu atp.

Práce s literaturou je vesměs adekvátní. Drobnou výtku bych měl k preferenci užívání on-line zdrojů. Myslím, že některé by bylo možné nahradit standardně vydanými publikacemi.

Závěrem konstatuji, že diplomant splnit zadání ve všech jeho bodech. Při celkovém hodnocení beru do úvahy fakt, že práce se systémem Mahout byla náročná. Nelze také pominout, že oblast strojového učení a zpracování přirozeného jazyka je teoreticky dosti obtížná. Dle mého názoru je naprosto zřejmé, že diplomant prokázal, že je schopen inženýrským způsobem řešit zadané problémy. Jeho práci doporučuji k obhajobě a hodnotím klasifikačním stupněm

“výborně”.



Ing. Miloslav Konopík, Ph.D.  
(vedoucí DP)

V Plzni 2. června 2015