

Posudek oponenta bakalářské práce

Autorka práce: Adam Mištera

Název práce: Full-textové vyhledávání s podporou porozumění textu dotazu

Obsah práce:

Práce se zabývá obohacením vyhledávacího dotazu pomocí metod sémantické reprezentace, integrací do Apache Solr a experimentům pro ověření funkčnosti metod. Obsah práce odpovídá zadání a pokrývá veškeré body zadání. Seznamuje nás s metodami sémantické reprezentace, full-textovým vyhledáváním v Solr a datovými kolekcemi pro evaluaci výsledků. Dále pak popisuje návrh a implementaci řešení, použité metriky a výsledky experimentů.

Formální úroveň:

Rozsah je v očekávaném rozmezí. **Logická struktura** má mírné nedostatky, například rozdělení do kapitol viz kapitoly 2.6 *Metriky* a 5.1 *Evaluační metriky*. Text obsahuje veškeré informace potřebné k porozumění problematice, jen ne vždy v pořadí, ve kterém by to vyhovovalo čtenáři. Popis kódu a seznam testovacích témat je uveden pouze v přílohách.

Stylistická stránka práce má prostor pro zlepšení (konzistence v termínech, překlepy, detailnější popisy pro lepší porozumění a reprodukovatelnost).

Velmi nešťastná je volba pozice popisů obrázků a tabulek (nad tabulkami a pod obrázky), což například na straně 35 může vést k záměně modelů. Také obrázky na stranách 39 a 41 by si nejspíše zasloužily zvolení stejného rozsahu na osách pro lepší porovnatelnost modelů.

Kvalita řešení a dosažených výsledků je nadprůměrná především rozsahem práce a množstvím experimentů. Kód je srozumitelný, dostatečně komentovaný a lze ho adaptovat pro podobné úlohy. Bohužel kód pro experimenty a trénování sémantických modelů na přiloženém DVD chybí, naštěstí jsem dostal přístup na git repozitář se zdrojovými kódy.

Práce s literaturou:

Uvedené zdroje jsou relevantní a dostatečné. Citační styl v Latex zkracuje autory prací do zkratky „et al.“, což se mi nezdá vhodné.

Dotazy k práci:

- 1) Jak vypadá kód pro trénování sémantických modelů?
- 2) Šlo použít stejnou množinu testovacích dat pro všechny experimenty?
- 3) Jaký je rozdíl mezi kosinovou podobností, kosinovou vzdáleností a kosinovou mírou?
- 4) Jak lze dosáhnout 100% úplnosti?

I přes výše zmíněné výtky převážně k formální úrovni textu, považuji práci za velmi dobrou. Zadání bylo splněno bez výhrad, proto navrhuji hodnocení známkou **velmi dobře** a práci **doporučuji k obhajobě**.



V Plzni 14. 5. 2019

Ing. Tomáš Hercig, Ph.D.