

Oponentský posudek Disertační práce

Ontology Learning and Information Extraction for the Semantic Web

Ing. Martina Kavalce

Obsah práce

Předložené práce se zabývá aktuální problematikou sémantického webu. Práce si klade za cíl prozkoumat extrakci informací, učení ontologií a sémantický web. Tyto tři oblasti slouží jako základ předkládaného výzkumu.

Práce podává ucelený přehled aktuálního stavu výzkumu v této oblasti. V úvodní části jsou uvedeny potřebné pojmy.

V další části je shrnut současný stav studované problematiky. V této části je uveden poměrně podrobný přehled metod pro extrakci informací založených na RDF.

Některé pojmy v úvodních kapitolách jsou popsány bez dalšího následného použití.

Jádrem práce je kapitola pátá, která obsahuje vlastní experimenty a jejich analýzu. Závěrečná šestá kapitola shrnuje dosažené výsledky a ukazuje na jistá omezení pro mapování skrytých relací do ontologií.

Na závěr jsou uvedeny experimenty a slovník termů.

Dotazy a připomínky

Práce obsahuje drobné nepřesnosti:

strana 17, vzorec na konci této stránky obsahuje symboly jenž nebyly definovány
strana 19, v textu jsem nenašel obrázek 2.3.3.

strana 50, zde se definuje $VCC(n_c, n_v)$ ale na straně 51 se používá $VCC(n)$

strana 55, definuje se zde podmnožina konceptů C^* ale značí se $C^* \in C$

strana 59, výsledky $AE(c_1 \vee c_2/v)$ jsou tabulce 5.1 v procentech ale v tabulce 5.2 jsou to reálná čísla.

Obrázky a diagramy jsou poměrně špatně čitelné.

Nejednotné používání odkazů na obrázky fig. nebo figure

K práci mám následující dotazy:

S výjimkou strany 53 z uvedeného textu nikde neplyne jaká je časová a prostorová složitost předkládaného řešení. Přesto, že jsou v práci uvedeny parametry stroje na kterém byly prováděny experimenty. Zajímalo by mě k jakým výsledkům autor dospěl.

Dále by mne zajímalo jak autor dospěl k hodnotě prahu pro VCC a CC . (strana 51). Předpokládám, že tato hodnota byla získána na základě experimentů a jejich následné analýzy.

Závěrečné hodnocení

Práce svědčí o autorově schopnosti jasně a srozumitelně formulovat tvrzení a velmi dobrém zvládnutí rozsáhlého teoretického aparátu. V práci jsou také obsaženy velmi dobře zpracované experimenty. K těmto experimentům byly použity texty Lonely Planet a SemCor.

Práce přináší nové poznatky.

Výsledky práce byly prezentovány na prestižních mezinárodních konferencích.

Celkovou úroveň práce lze hodnotit velmi dobře.

Závěr

Práci doporučuji k obhajobě před příslušnou komisí.

V Ostravě dne 1. září 2006
Prof. RNDr. Václav Snášel CSc,
Katedra informatiky
Fakulta elektrotechniky a informatiky
VŠB-Technická Univerzita Ostrava