

41₁₀ – Aby byla anotace A sémantická nemůže být vždy konzistentní s tím, jak lidé znalosti a informace chápou. Stačila by existence jediného, kdo ji tak nechápe a už bychom A mohli zahodit.

41₆ – „širokost“ publika není pro standardnost rozhodující. To publikum by muselo být omezeno na jednu komunitu resp. ty, kteří využívají jednu aplikační doménu.

42¹ – úvahy o ideální anotaci nejsou nikterak užitečné.

42₁₀ – naopak užitečná je úvaha o vztahu lingvistické a sémantické anotace. Přístupy v tomto směru byly již popsány v 80. letech – viz práce Materny, Duží a dalších založené na pojmu atribut a teorii konstrukcí.

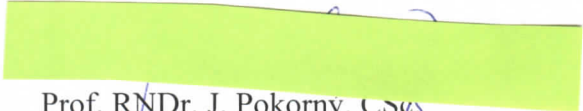
43 – odst. 3.5.2 – uvedené nejednoznačnosti používají rádi lingvisti pro své argumentace. V daném kontextu však působí spíše obskurně.

56 – odstavce 4.2 – 4.4 jsou stručnou sbírkou informací bez jasného vztahu k cíli práce. Co jsou výsledky? Zřejmě ne pouze frekvenční analýza zpráv.

81 – celý odst. 6.3.1. Zmínky o software a nemožnosti jeho provozování jsou velmi matoucí. Jaké licence, jací autoři práce? Jaké prodávání software? Zatím není nikde popsán důvod, proč by si měl někdo nějaký software kupovat.

Závěr: Práce, jakkoliv jsou některé její části zajímavé a pro začátečníka v dané oblasti i přínosné, je svým zpracováním na hranici uznatelnosti za diplomovou. Bude na autorovi, aby při její obhajobě prokázal existenci výsledků, které se v práci tak obtížně hledají.

V Praze dne 22. 8. 2007



Prof. RNDr. J. Pokorný, CSc.
KSI MFF UK

Oponentský posudek na diplomovou práci Jana Dědka

Sémantická anotace dat z webovských zdrojů

Předmětem práce je téma, které je velmi aktuální a současně také velmi obtížné. Cílem bylo jednak mapovat různé přístupy, metody, projekty týkající se tématu, dalším bylo zaměřit se na nějakou doménu a tam využít nabyté znalosti resp. nástroje ke smysluplnému experimentu. Zdá se, že autorovi se spíše zdařilo přiblížit se dosažení prvního cíle.

Anotovat různé přístupy na jedné straně nebylo zas tak pracné, stačilo přečíst a zpracovat odpovídající množství literatury. To autor provedl a docela čtivým způsobem popsal state-of-the-art dané oblasti. Někdy poněkud stručně, někdy poněkud v nelogické struktuře (např. v 3.3.5 – hovoří se o dvou kategoriích metod, nicméně podnadpisy jsou tři, navíc druhý má zcela nekompatibilní název), nicméně na slušné informativní úrovni.

Jakkoliv jsou kapitoly 1-3 celkem zajímavé, v dalším textu se autor pouští do oblastí, kde je stále stručnější a nakonec, v kapitole 6, se dostává k textu, z něhož již vůbec není jasné, do vlastně bylo obsahem experimentu, co je uděláno a co je pouze ve stádiu vizí (viz např. použití WordNetu).

Škoda, že autor nedal práci přečíst někomu s kladnějším vztahem k češtině. Jistě by tak odstranil některé systematické chyby, zvláště pak v koncovkách sloves v minulém čase a množném čísle (12 – „že jsme narazily“, 17 – „softwaroví agenti se vyznali“, 23 – „prvky odpovídali“, 32 – „nástroje vyhodnocovali“), ale i jinde (14 – softwarový agenti“).

Práci také neprospívá spousta anglicismů, někdy i v místech, kde je několik českých ekvivalentů (11 - web-stránka, 15 – Data Miningu, vše na obrázku 2.1, 25- triple, namespace prefix, 28 – reasoning, 31 – pointery, 59 – nebo expresivita – to je přece vyjadřovací síla).

Několik drobných poznámek, kritických připomínek či nejasností:

21⁴ – co je destrukční jazyk?

33 – 3.3.2 – pozor slovo wrapper se používá v mnoha jiných souvislostech v informatice. Bylo by dobré ho pořádně vysvětlit.

37¹³⁻¹⁶ – možnost definovat typ atributů není jejich vlastnost. Tou vlastností je přímo ten typ. Co je vlastnost? Funkce, který přiřazuje něčemu něco. Proto se také atributy nepřepočítávají, ale přepočítávají se jejich hodnoty.

39 – co znamená věta, že se musí Lixto ručně naprogramovat na každou stránku?

41 – Úvahy v odst. 3.5.1-3.5.1 jsou spíše filosofického než teoretického rázu.