Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

☐ posudek vedoucího ☒ posudek oponenta

Autor/ka: Ján Pecsők

Název práce: Vzájemné odkazování slov v textu

Studijní program a obor: Informatika

Rok odevzdání: 2011

Jméno a tituly vedoucího/oponenta: Mgr. Michal Novák

Pracoviště: ÚFAL MFF UK

	excelentní	**odpovídající**	**slabší**	**nevyhovující**
Náročnost zadaného tématu		x		
Míra splnění zadání		x	x	
Rozsah práce		x		
Struktura textové části práce			x	x
Analýza		x	x	
Vývojová dokumentace			x	x
Uživatelská dokumentace		x		
Jazyková a typografická úroveň				x
Návrh a design implementace		x		
Kvalita zpracování softwarové části			x	
Stabilita aplikace		x		

Nejvýznamnější klady:

- Na práci sa mi páči nevšedný a flexibilný prístup pri rozpoznávaní koreferenčných vzťahov. Autorov algoritmus rozpoznávania nie je fixovaný na hľadanie vhodného antecedentu (prvý člen) k anaforu (druhý člen vzťahu), ako je pri tejto úlohe bežné. Naopak, schéma pravidiel je z tohto pohľadu univerzálna a umožňuje hľadať aj anafor k antecedentu. Takisto umožňuje hľadať v nasledujúcom texte, čím sa dajú prípadne rozpoznať kataforické vzťahy.

Nejzávažnější nedostatky:

- Formálna stránka je nevyhovujúca – gramatické a typografické chyby, chybné odkazy k obrázkom a literatúre, zlá kvalita obrázkov, rôzne formátovanie textov rovnakého typu a naopak rovnaké formátovanie textov rôzneho typu, nekonzistentný formát bibliografie atď. Po štylistickej stránke je text tiež slabší. Celkovo mám pocit, že práca nebola ani raz revidovaná.
- Rušivým dojmom pôsobí umelé naťahovanie textu, napr. nezaujímavými screenshotmi alebo niekoľko strán dlhými doslovnými citáciami zo stránok PDT (str. 11) a slovenskej Wikipedie (str. 24-26), čo navyše hraničí s dobrými mravmi pri písaní odbornej práce. Naopak, práca by si zaslúžila hlbšiu analýzu problému a zdôvodnenie výberu zvoleného riešenia z množstva iných možných riešení. Obrovským nedostatkom je absencia diskusie a záveru, práca končí naozaj nečakane. Takisto vývojová dokumentácia by mala byť obsiahlejšia a súčasťou textu práce, neprehľadný flowchart a heslovitý JavaDoc na CD nestačí.
- Po obsahovej stránke mám výhradu hlavne k použitej metrike. Zanedbanie druhého člena páru umelo zlepšuje výsledok, napr. nech slová A,B,C,D vytvárajú anotované páry (A,B) a (C,D). Ak však systém nesprávne vytvorí páry (A,C) a (B,D), úspešnosť na základe navrhnutej metriky je 100%. Otázka vhodnej metriky na evaluáciu koreferencie je veľmi diskutovaná, najpoužívanejšie sú pairwise F-score, MUC, B_3 a ϕ_3-CEAF.
- To, že aplikácie vyšetruje koreferenčné reťazce po dvojiciach slov je bežné a v poriadku. Avšak pri vizualizácii by sa mali jednou farbou a jedným indexom zobraziť všetky slová patriace do rovnakého reťazca – tranzitívneho uzáveru. V skutočnosti každé slovo dostane toľko indexov, v koľkých pároch vystupuje. To robí výstup veľmi neprehľadným a nájsť v zobrazenom texte celé koreferenčné reťazce si vyžaduje netriviálny manuálny postprocessing.

Další poznámky:

	výborně	**velmi dobře**	**dobře**	**neprospěl/a**
Návrh známky			x	

Datum: 25. 8. 2011

Podpis: