

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucích posudek oponenta

Autor/ka: Jakub Záhumenský

Název práce: Hľadanie odpovede v odpovediach

Studijní program a obor: Informatika, obecná informatika

Rok odevzdání: 2011

Jméno a tituly vedoucího/oponenta: RNDr. Ondřej Bojar, Ph.D.

Pracoviště: ÚFAL MFF UK

	e x c e l e n t n í	o d p o v í d a j í c í	s l a b š í	n e v y h o v u j í c í
Náročnost zadaného tématu		X		
Míra splnění zadání		X		
Rozsah práce		X		
Struktura textové části práce		X		
Analýza	X	X		
Vývojová dokumentace		X		
Uživatelská dokumentace		X		
Jazyková a typografická úroveň		X		
Návrh a design implementace		X		
Kvalita zpracování softwarové části		X		
Stabilita aplikace		X		

Nejvýznamnější klady:

Ucelená práce, vyzkoušeno několik metod hledání nejpodobnějších otázek, základní evaluace metod.

Nejzávažnější nedostatky:

Občas mírně matoucí pojmenování: fiktivní a reálný mód rozhovoru (dalo mi práci motivaci názvů pochopit), adaptabilní stoplist (nijak se sám neadaptuje, uživatel může zvolit práh). Vyloženě zavádějící je pak název “5.3.2. Morfologická analýza”, která používá lemma+značku, tj. de facto formu slova. “Frekvenční analýza” naopak užívá pouze lemma, a opírá se tedy o morfologickou analýzu textu.

Nepřesnosti ve formálním popisu, Lemma 1 na str. 24: “najpravejšie korene” – strom má jeden kořen, tak jakýpak nejpravější? Navíc “zakorenený strom” ještě neříká, že uzly ve stromě jsou uspořádány. Mělo by se tedy říci “uspořádané kořenové stromy”. V lemmatu je dále zmíněno “usporiadané lesy” – nebo lesy uspořádaných stromů? Nebo skutečně chcete říci, že les má stromy (lineárně?) uspořádané?

Nejasnost v evaluaci: Jaký je vztah mezi listováním mezi 3 nabídnutými odpověďmi a potvrzením “byl jsem spokojen”. Spokojen s alespoň jednou z těch tří odpovědí? Nebo spokojen s tou první nabídnutou? Druhým problémem evaluace je, že nezohledňuje, jestli vůbec v korpusu vhodná odpověď byla, pro ten účel by se správnou odpověď musel pokusit najít člověk.

Další poznámky:

Je použit opravdu zastaralý nástroj tool_chain, pro hladké nasazení na jiné texty bude překážkou kódování ISO-8859-2, které nástroj vyžaduje, ale dnešní české texty často překračují (např. typografickými uvozovkami).

Občasné chyby v psaní mezer kolem závorek a další interpunkce, jinak prakticky bez překlepů.

Zkratka “cf” v tabulce na Obr. 4 není zavedena, je však zřejmé, že jde o “collection frequency”.

Jak je získán slovník synonym, bylo by jej možné získat strojově?

Proč se pro podobnost používá skalární součin a nikoli kosinová podobnost, tj. normovaný skalární součin?

Morfologicky nerozpoznaná slova (značka X@...) jsou v některých variantách analýzy zcela ignorována. Lepší by bylo použít alespoň slovní formu nebo velmi hrubou lematizaci: první čtyři písmenka slova.

	v ý b o r n ě	v e l m i d o b ř e	d o b ř e	n e p r o s p ě l / a
Návrh známky	X			

Datum:13.6.2011

Podpis:

Poučení k formuláři pro hodnocení infromatických bakalářských prací

Tento formulář je určen pro hodnocení vedoucího i oponenta bakalářské práce, která má formu softwarového projektu. Bakalářské projekty jiných typů (teoretická práce, srovnávací studie apod.) budou hodnoceny pomocí standardních textových posudků.

Jednotlivá políčka vyplňte nejlépe elektronicky (lze případně i ručně), je možné zaškrtnout i dvě sousední políčka (např. pro hodnocení typu 'něco mezi odpovídající a slabší'), a to i u návrhu výsledné známky. Pokud některá položka nemá vzhledem k práci smysl (např. stabilita aplikace u práce bez vlastní implementace), položku nevyplňujte. Výsledná navrhovaná známka nemusí být žádným 'průměrem' hodnocení jednotlivých kritérií. Pokud některé položky hodnotíte jako slabší nebo nevyhovující, v sekci Nejzávažnější nedostatky popište důvody vašeho hodnocení a zjištěné nedostatky.

Výklad stupňů hodnocení:

- excelentní znatelně lepší/rozsáhlejší/dokonalejší než je pro Bc práci požadováno
- odpovídající přiměřené Bc práci, student splnil to, co měl
- slabší výhrady ke kvalitě, rozsahu, hloubce nebo zpracování
- nevyhovující neodpovídá požadavkům na Bc práci, práce nemá být obhájena

Vyplněné a ručně podepsané (i v případě elektronického vyplňování) hodnocení odevzdejte na sekretariát KSI, elektronickou verzi pošlete na sekretariat@ksi.ms.mff.cuni.cz. Pokud máte emailový kontakt na autora práce, pošlete posudek i jemu.