

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucího posudek oponenta

Autor/ka: Lenka Smejkalová

Název práce: Vizualizace podobnosti dokumentů

Studijní program a obor: Informatika, Programování

Rok odevzdání: 2007

Jméno a tituly vedoucího/opponenta: Mgr. Pavel Schlesinger

Pracoviště: ÚFAL MFF UK

	ex cel en tní	od po ví- da jic í	s la bš í	ne vy- ho vu jící
Náročnost zadaného tématu		X		
Míra splnění zadání	X			
Struktura textové části práce	X			
Jazyková a typografická úroveň		X		
Analýza		X		
Vývojová dokumentace	X			
Uživatelská dokumentace	X			
Kvalita zpracování softwarové části	X			
Stabilita aplikace	X			

Nejvýznamnější klady:

Vynikající, dobře čitelná práce, ve které autorka v rozsahu požadovaném od bakalářské práce uceleně zpracovala a čtenáři představila jeden z principů, jak porovnávat dokumenty, eventuálně vyhledávat duplicity. Kladně oceňuji zejména množství vhodně zvolených doplňujících obrázků a příkladů v textu. V tomto ohledu by odevzdaná práce mohla být vzorem pro ostatní.

Téma vizualizace podobnosti textu je v tomto případě uživateli přiblíženo aplikací Dotplot (grafická verze i verze ovladatelná z příkazové řádky). Vyzdvihuji snadno a intuitivně ovladatelnou grafickou verzi nabízející okamžité experimentování, zobrazení výsledků a snadnou změnu parametrů.

Nejzávažnější nedostatky:

Další poznámky:

Zvolené téma nabízí řadu dalších vylepšení; některé z nich autorka sama zmiňuje v podkapitole 5.1. Já sám bych navrhol zajímavou orientaci na analýzu výsledných obrázků a podrobnější studium možných vlastností textu a relací mezi dokumenty, které lze popsanou metodou detekovat. Rovněž sama podobnost (similarity) se dá měřit větším počtem statistických měr, než bylo zde použito.

Čitelnosti textu by ještě víc pomohlo, kdyby příklady v textu byly VŽDY doplněny odkazem na příslušný obrázek. Také je lepší se vyvarovat užívání odborného anglického termínu bez jeho předchozího vysvětlení nebo uvedení alespoň českého ekvivalentu v závorce (jako na str. 7 self-similarity v jinak “české větě”). Tyto nedostatky jsou ovšem z hlediska klasifikace nepodstatné.

Autorka vyvíjela a testovala aplikaci pro Linux; mě se podařilo aplikaci bez problémů zkompileovat a provozovat i pod Mac OS X 10.4.10.

	vý bo rn ě	ve l- mi do bř e	do bř e	ne- pr os pěl /a
Návrh známky	X			

Datum: 4. září 2007

Podpis:
