

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucího

Autor/ka: Jan Hajič

Název práce: Popularita osob automaticky

Studijní program a obor: Informatika, obor 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	X		
Struktura textové části práce	X	X		
Analýza	X			
Vývojová dokumentace		X		
Uživatelská dokumentace		X		
Jazyková a typografická úroveň	X	X		
Návrh a design implementace		X		
Kvalita zpracování softwarové části		X		
Stabilita aplikace		X		

Nejvýznamnější klady:

- samostatnost při práci, široký záběr, lingvistická i technická analýza
- srovnání domény novinových článků s doménou recenzí (nad rámec zadání), což doložilo, že problém rozpoznat polaritu do značné míry spočívá v datech, nikoli metodě
- práce představuje velmi solidní první krok k „analýze smýšlení“ (sentiment analysis) v češtině; z dlouhodobého hlediska je nejcennější návrh a následná analýza problémů ruční anotace

Nejzávažnější nedostatky:

- sekce výsledků je velmi hutná, podrobnější komentář k tabulkám by čtení usnadil.
Autor potřebné rozvahy nad naměřenými výsledky evidentně provádí, mohl však čtenáře šetřit a přímo mu je předestřít. Lze uvést dva příklady:
 - Neúspěch na člancích (oddíl 4.6.2) je jasně patrný z nízké přesnosti a úplnosti, ale „accuracy“ vychází 70 %, protože zahrnuje i velmi časté případy, kdy počítač korektně neprovedl nic. Čtenáře poměrně vysoké číslo bez komentáře může zmást.
 - Vyhodnocování na trénovacích datech (ve srovnání s testovacími daty) má dobré důvody, čtenář však může snadno tento fakt přehlédnout, není-li v komentáři zdůrazněn.

Další poznámky:

- hlavní algoritmus klasifikace by bylo dobré popsat i slovně, nikoli jen vzorečky
- jak často se uplatnilo hrubé a zjemněné řešení negace? To by mohlo částečně vysvětlit rozdíl mezi výsledky těchto metod.
- v práci je velmi málo překlepů, automatickou kontrolou překlepů však neprošla

	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: 21.8.2011

Podpis: