

Posudek bakalářské práce

předložené na Matematicko-fyzikální fakultě
Univerzity Karlovy v Praze

posudek vedoucího posudek oponenta

Autor/ka: Pavel Češka

Název práce: Segmentace textu

Studijní program a obor: Informatika, obecná informatika

Rok odevzdání: 2006

Jméno a tituly vedoucího/opponenta: Mgr. Pavel Pecina

Pracoviště: ÚFAL MFF UK

Náročnost zadaného tématu
Míra splnění zadání
Struktura textové části práce
Jazyková a typografická úroveň
Analýza
Vývojová dokumentace
Uživatelská dokumentace
Kvalita zpracování softwarové části
Stabilita aplikace

e	o	s	n
x	d	l	e
c	p	a	v
e	o	b	y
l	v	š	h
e	í	í	o
n	d		v
t	a		u
n	j		j
í	í		í
	e		e
	í		í
	X		
X			
X			
X			
X			
		X	
		X	
X			
X			

Nejvýznamnější klady:

Výsledek této práce je zřejmě zdaleka nejlepší nástroj pro tokenizaci, segmentaci vět a identifikaci jazyka, který je pro češtinu k dispozici. Kvalitu ostatních, většinou ad-hoc řešení převyšuje v množství rozpoznávaných zkratk, použitou statisticko-klasifikační metodou detekce konců vět a analýzou několika různých přístupů (a jejich vlastní modifikací) k identifikaci jazyka. Úspěšnost segmentace a identifikace vět byla navíc nezávisle ověřena a testována na reálných datech, což není u prací tohoto druhu běžné.

Nejzávažnější nedostatky:

Vyvojová dokumentace je pouze ve formě komentovaných zdrojových kódů, protože se ale nepředpokládají výraznější zásahy do kódu není tento nedostatek příliš závažný. Uživatelská dokumentace je řešena formou on-line nápovědy a readme souboru, která je ale také zcela dostatečná a umožňuje plné využití aplikace. Nápověda je navíc v anglickém jazyce a tudíž bude program jednoduše použitelný i pro daleko širší okruh pracovníků.

Další poznámky:

Výsledek této práce je natolik kvalitní, že se předpokládá jeho plné nasazení na Ústavu formální a aplikované lingvistiky na MFF UK, případně i na jiných pracovištích.

	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: 15.6. 2006

Podpis:

