

Posudek oponenta diplomové práce

Jméno a příjmení autora posudku: Doc. RNDr. Vladislav Kuboň, Ph.D.

Jméno a příjmení autora práce: Bc. Petr Fanta

Název práce: Kontextově závislý slovník pro překladatele

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Předkládaná diplomová práce se zabývá problémem vytváření pomocných slovníků, které by doplnily informace, jež má k dispozici překladatel, na bázi automatického zpracování kontextu a vygenerování relevantních slovníkových hesel s vysvětlením významu jednotlivých pojmů. Jedná se o úkol obtížný, skrývající v sobě celou řadu poměrně složitých dílčích úkolů.

Diplomová práce se skládá z osmi kapitol. První kapitola uvádí hlavní cíl práce a popisuje její strukturu, druhá kapitola analyzuje daný problém a seznamuje čtenáře s odborným kontextem. Třetí kapitola představuje hlavní datové zdroje, které jsou v práci využity, čtvrtá analyzuje problém automatického vytváření speciálního anglicko-českého slovníku. Implementaci systému je věnována kapitola 5 a poměrně rozsáhlá kapitola 6, tvořící cca 20% celé práce, se podrobně zabývá vyhodnocením úspěšnosti systému. Kapitola 7 pak nastiňuje další možné směry vývoje a poslední, osmá, kapitola pak stručně shrnuje dosažené výsledky. Práce je vhodně doplněna přílohami a rozsáhlou bibliografií, kterou autor v práci správně cituje.

Po přečtení celé práce čtenář nabyde několika dojmů. Ten první je potvrzení faktu, že vytváření kontextově závislých slovníků je velmi komplikovaný úkol, jehož úspěšné zvládnutí předpokládá důkladné seznámení se s celou řadou nástrojů pro zpracování přirozeného jazyka a jejich následnou kombinaci do jednoho rozsáhlého celku. Druhým dojmem je přesvědčení, že autor velmi dobře rozumí tomu, co dělá, dokáže si poradit i s takto obtížným úkolem a práce nevykazuje žádná závažná opomenutí nebo přílišné zjednodušování. Třetím dojmem, který je podpořen zejména některými výsledky, podrobně představenými v šesté kapitole, je pochybnost o užitečnosti výstupů systému pro překladatele. To ovšem není autorova chyba, ten pouze důkladně splnil zadání práce.

Určité pochybnosti vyvolávají už motivační příklady, na kterých se autor snaží demonstrovat potřebnost pomocných slovníků, které budou nejen překládat jednotlivá slova nebo víceslovné pojmenované entity, ale také vysvětlovat jejich význam. Tato pochybnost spočívá v tom, že pokud většinu předkládaných příkladů přeložíme bez těchto dodatečných informací, dostaneme v podstatě stejně dobrý nebo špatný překlad jako s nimi. K tomu, abychom přeložili věty se jmény Omar al Shishani nebo John Kerry, nepotřebujeme vědět, kdo tito pánové jsou, stačí nám pouze schopnost dokázat tato jména identifikovat v textu a eventuálně (v Omarově případě) je mechanicky přepsat podle zvyklostí cílového jazyka. Na druhou stranu u méně obvyklých termínů či zkratk z příkladů na str. 6 (SMM, mortar, Grad) může být vysvětlení pro amatérské překladatele užitečné. Pro čtenáře tedy nutně vyvstává otázka, zda by nebylo vhodnější množství poskytovaných informací nebo i počet hesel, které v každé vstupní větě systém hledá, více omezit.

Toto přesvědčení posiluje i metodicky správně provedené vyhodnocení užitečnosti systému v kapitole 6.6., ze kterého vychází, že pouze u 39%, tedy zhruba u dvou pětin hesel, hodnotitelé uvedli, že "existuje nějaká informace, která by jim pomohla s překladem". Přestože musíme brát v úvahu subjektivitu tohoto hodnocení, protože jednotliví překladatelé samozřejmě disponují různou mírou znalostí nutných k porozumění textu, je toto číslo poměrně nízké. Je škoda, že se jej autor nepokusil nějak zvýšit a výzkum zaměřený na tuto oblast uvádí v sedmé kapitole mezi mnoha jinými nápady na další vývoj systému.

Po obsahové stránce je celou práci možné hodnotit jako velmi dobrou, autor přehledně popisuje jednotlivá rozhodnutí a kroky při analýze a vývoji systému, je si vědom důsledků těchto rozhodnutí a seznamuje s nimi i čtenáře. Je možné konstatovat, že se autor velmi dobře vyrovnal jak s vývojem a implementací daného systému, tak i s jeho popisem v předkládané práci. Textová část práce umožňuje čtenáři udělat si dobrý obrázek o daném problému i o jeho řešení, přestože je práce svým rozsahem spíše na dolní hranici přijatelnosti. To však u implementačně zaměřené práce je možné tolerovat.

Příznivý dojem z práce poněkud sráží řada překlepů a dokonce i výskyt hrubých gramatických chyb (např. „problémy se objevili“ na str.6, či „slovníky obsahovali“ na straně 45).

Doporučení k obhajobě:

Z výše uvedených důvodů práci *doporučuji* k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	ANO <input type="checkbox"/>
---	------------------------------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

--

V Praze dne: 20.1.2017

Podpis:**

** nehodící se škrtněte (vymažte)*

*** do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*