

Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Karolína Burešová
Název práce Text Simplification in Czech
Rok odevzdání 2017
Studijní program Informatika **Studijní obor** Matematická lingvistika

Autor posudku RNDr. Pavel Pecina Ph.D. **Role** vedoucí
Pracoviště Ústav formální a aplikované lingvistiky

Text posudku:

Diplomová práce Karolíny Burešové se zabývá zjednodušováním textu v češtině. Tuto úlohu lze definovat jako přeformulování vstupního textu tak, aby výsledek byl pro čtenáře jednodušší a tím snadněji pochopitelný. Tato úloha může mít různá uplatnění, např. pro usnadnění pochopení obsahu textu nerodilými mluvčími nebo jedinci s poruchami porozumění textu. Úlohu lze řešit na různých úrovních: lexikální, syntaktické, ale i pragmatické. Autorka se zabývá úrovní nejnižší, tedy nahrazováním slov jejich (jednoduššími) synonymy.

Práce je psaná anglicky, celkem na 85 stranách, rozdělená do 8 kapitol, opatřená bibliografií a popisem přílohy, kterou tvoří 1 ks CD s implementací navržených metod a daty, která byla použita v experimentální části. První dvě kapitoly jsou úvodní, seznamují čtenáře s problematikou, existujícími přístupy a specifiky, která přináší řešení této úlohy pro češtinu. Třetí kapitola podrobně diskutuje různé aspekty a metody zjednodušování textu, zejména potom metody odhadování „čitelnosti“ a „srozumitelnosti textu“. Ve čtvrté kapitole autorka popisuje experiment, ve kterém zjišťovala, jak úroveň srozumitelnosti/čitelnosti textu vnímají lidé. V páté kapitole popisuje navrženou a implementovanou metodu a šesté kapitole výsledky provedených experimentů. Sedmá kapitola popisuje některé implementační detaily. Osmá kapitola uzavírá práci diskusí výsledků a možný směrů dalšího výzkumu.

Práci lze charakterizovat jako výzkumně-experimentální a částečně implementační s tím, že v experimentech byly použity existující nástroje (např. MorphoDiTa pro morfologické značkování a generování), ale stěžejní části použitých algoritmů autorka implementovala sama.

Celkově je práce psaná čtivě, bez rušivých gramatických chyb, s nepříliš častými překlepy. Místy je ovšem text zbytečně „upovídáný“ a nejde přímo k věci (např. „My supervisor suggested that [...]. I agreed this was an interesting idea and decided to explore it further.“). Autorčin styl je charakteristický používáním dlouhých vět a krátkých odstavců, což občas vede k situaci, že odstavec tvoří jedna věta. Porozumění textu to nezhoršuje, ale při čtení to může působit rušivě.

Celkově je struktura textu dobrá se všemi náležitostmi odborné práce tohoto typu (popis metod, experimentů, výsledků a jejich diskuse).

Experimentální stránka práce je nadprůměrná. Autorka si připravila svoje vlastní trénovací a testovací data, provedla experimenty s lidským vnímáním srozumitelnosti a čitelnosti a řadu experimentů s automatickými metodami. Inspiraci hledala v metodách nevržených pro jiné jazyky (angličtinu), ty ale musely být upraveny tak, aby byly použitelné na jazyk s výrazně bohatší morfologií a jinými jazykovými zdroji (slovníky). Výsledky experimentů jsou dobře zpracovány, prezentovány a diskutovány. Chybějí snad jen rozsáhlejší příklady výstupů jednotlivých (vybraných) experimentů. Vyzdvihl bych také velmi dobře zpracovanou část teoretickou, které poskytuje velmi dobrý základ pro další práci na této úloze.

Práce Karolíny Burešové je prvním počinem v zjednodušování textu v češtině. Byla v ní prozkoumána oblast lexikálního zjednodušování (s několika omezeními, jasně zmíněnými v textu), možnosti další práce jsou shrnuty v závěrečné kapitole a definují tak směr dalšího výzkumu, kterým se chce autorka zabývat v rámci doktorského studia.

Předkládaná práce splňuje požadavky kladené na diplomovou práci na MFF UK. Její cíle byly splněny.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 1. 6. 2017

Podpis: