

KATEDRA OBECNÉ LINGVISTIKY

Posudek na magisterskou diplomovou práci

Autor: Vladimír Matlach

Studijní obor: obecná lingvistika a teorie komunikace

Název práce: Kvantitativně lingvistický software

Vedoucí práce: Radek Čech

Hodnotící kritéria	Hodnocení
	1–4
Soulad názvu práce s obsahem, schopnost formulovat cíl a závěry práce	1
Teoretická podkladová koncepce, metodologie, terminologie	1
Schopnost analýzy excerpovaného materiálu/předmětu výzkumu a následné syntézy	1
Interpretace a argumentace	1
Sekundární literatura, tvůrčí aplikace, schopnost kritického komentáře	1
Struktura práce, výstavba textu, adekvátnost použitých stylistických prostředků, dodržení ortografické a gramatické normy	1,5
Bibliografický a poznámkový aparát, jednotný způsob citací	2
Výsledný průměr	1,21

Hodnocení práce

Cílem dipl. práce bylo vytvoření softwaru pro kvantitativnělingvistickou analýzu textů. Tento cíl byl splněn na vynikající úrovni – V. Matlachovi se podařilo vytvořit software, který nabízí celou řadu funkcí: od tokenizace a lemmatizace textů, přes relativně složité výpočty indexů vyjadřujících nejrozumnější kvantitativní charakteristiky, až po grafické znázornění a uložení výsledků. V softwaru jsou navíc implementovány funkce, které u vybraných indexů umožňují statisticky testovat významnost rozdílu měřených hodnot. O kvalitě softwaru jistě svědčí i to, že nachází ohlas mezi kvantitativními lingvisty – srov. Kubát, M., Matlach, V., Čech, R. (2014). Announcement: Quantitative Index Text Analyser (QUITA). Glottometrics, 27, 91-92, a prezentaci softwaru na mezinárodní konferenci „QUALICO 2014 20th anniversary of IQLA and Journal of Quantitative Linguistics (JQL) Olomouc (Czech Republic), May 29 - June 1, 2014“.

V textové části diplomové práce autor velmi přehledně a srozumitelně představuje základní vlastnosti softwaru a ilustruje je na konkrétních případových studiích. V této části prokazuje také schopnost lingvistické interpretace – oceňuji zejména to, že se snaží ilustrativní příklady prezentovat ve formě lingvistických problémů, které mají formu empiricky testovatelných hypotéz. Dále oceňuji, že jsou otevřeně prezentovány nedostatky a limity softwaru (např. lemmatizace) a že je celý projekt navržen takovým způsobem, že může být dále upravován a vylepšován.

Nedostatky práce

Zásadní nedostatky se v práci neobjevují, jde jen o dílčí chyby:

s. 11-12

Porovnání kódu napsaného v R a toho, který je použit v QUITA, je neadekvátní. V příkladu výpočtu v R je totiž zapsán i poměrně složitý výpočet rozptylu, kdežto v QUITA nikoliv.

s. 78

Seznam literatury by měl být seřazen buď abecedně, nebo očíslován.

V textu se občas vyskytují pravopisné chyby a překlipy.

Hodnocení písemné verze práce – návrh klasifikačního stupně:

výborně

Náměty pro obhajobu

Bylo by možné upravit funkci vytvoření randomizovaného textu tak, aby výsledný „text“ obsahoval i náhodně vytvořené „věty“, případně odstavce?

Je možné do QUITA implementovat funkce, které by umožnily i syntaktickou analýzu textu (např. přiřazení větných členů, sémantických rolí apod.)?