

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Ján Faryad	
Název práce	Rozpoznávání koreference pro Universal Dependencies	
Rok odevzdání	2018	
Studijní program	Informatika	
Studijní obor	Obecná informatika	
Autor posudku	Mgr. Michal Novák	Vedoucí
Pracoviště	Ústav formální a aplikované lingvistiky	

K celé práci

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	X			
Splnění zadání	X			
Rozsah práce <small>... textová i implementační část, zohlednění náročnosti</small>	X			
<p>Je nutné zdůraznit, že úloha a její řešení byly obtížné v tom, že po řešiteli vyžadovaly osvojení netriviálního množství nových znalostí a dovedností. Lidé pracující v oblasti IT tomuto musí čelit neustále a autor práce se s tím popasoval úspěšně. Z pohledu zpracování přirozeného jazyka musel nastudovat základy lingvistického popisu koreference a jejich typů a rovněž si prakticky vyzkoušel práci s textovými daty v různých stylech a formátech anotace, jejichž heterogenita je v této oblasti poměrně častým jevem. Co se týče inženýrských dovedností, autor se musel jako uživatel obeznámit s nástrojem UDPipe a coby programátor se musel zorientovat v architektuře frameworku Udapi, do kterého svoje moduly pro rozpoznávání koreference implementoval. Nově nabytí zkušenosti s programovacím jazykem Python, systémem Git na správu verzí a základními metodami strojového učení.</p> <p>Požadavky, které si autor vytyčil v zadání, splnil bez výjimky, v některých částech dokonce šel nad jejich rámec. Například navrhl a implementoval způsob rozpoznávání koreference vypuštěných podmětů v češtině, které jsou pro koreferenční vztahy naprosto zásadní, protože kvůli jejich absenci v podobě specifického povrchového výrazu není vůbec jednoduché je automaticky zpracovávat.</p>				

Textová část práce

lepší OK horší nevyhovuje

	lepší	OK	horší	nevyhovuje
Formální úprava <small>... jazyková úroveň, typografická úroveň, citace</small>		X		
Struktura textu <small>... kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</small>	X			
Analýza	X			
Vývojová dokumentace	X			
Uživatelská dokumentace	X			

Text splňuje požadavky kladené na bakalářskou práci svou strukturou, jazykovou, typografickou i stylistickou úpravou. K menším prohrškům vůči odbornému stylu bych zařadil použití familiárního termínu "dědeček"(str. 43) k označení prapředka v syntaktickém stromě. Pochopení trochu stěžuje absence použití zavedené terminologie, konkrétně u termínů "anafora"(lingvisticky správně by mělo být "anafor") a "antecedent", které odpovídají odkazujícímu a odkazovanému výrazu ve vztahu koreference. Termín "anafora"autor po zavedení dále v textu nepoužije vůbec a termín "antecedent"jen dvakrát. Anafor je tak občas označován jako zájmeno nebo zájmenný uzel (str. 40), což je matoucí, jelikož zájmenem může být jak anafor, tak i antecedent.

Implementační část práce

lepší OK horší nevyhovuje

Kvalita návrhu	<i>... architektura, struktury a algoritmy, použité technologie</i>	X			
Kvalita zpracování	<i>... jmenné konvence, formátování, komentáře, testování</i>	X			
Stabilita implementace			X		

System je dobře navržen s důrazem na jazykovou univerzálnost. Instalace a použití proběhlo téměř bez problémů. Snadno odhalitelným problémem byla chybějící specifikace proměnné SHELL v makefilech a volání Bash skriptů příkazem "sh". Může se tak spustit jiný shell, kde nejsou některé příkazy definovány (např. příkaz "source").

Celkové hodnocení Výborně
Práci navrhuji na zvláštní ocenění Ne

Datum

Podpis