

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Kateřina Nová
Název práce Analysis and visualization of OCR output
Rok odevzdání 2022
Studijní program Informatika
Specializace Obecná informatika

Autor posudku Barbora Vidová Hladká
Pracoviště ÚFAL MFF UK

Role Vedoucí

Prosím vyplňte hodnocení křížkem u každého kritéria. Hodnocení *OK* označuje práci, která kritérium vhodným způsobem splňuje. Hodnocení *lepší* a *horší* označují splnění nad a pod rámec obvyklý pro bakalářskou práci, hodnocení *nevyhovuje* označuje práci, která by neměla být obhájena. Hodnocení v případě potřeby doplňte komentářem. Komentář prosím doplňte všude, kde je hodnocení jiné než *OK*.

K celé práci

	lepší	OK	horší	nevyhovuje
Obtížnost zadání	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Rozsah práce ... <i>textová i implementační část, zohlednění náročnosti</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Cílem bakalářské práce bylo provést statistickou analýzu výstupu vybraného systému OCR a vhodně ji vizualizovat. Cíl zahrnoval splnění čtyř dílčích cílů:</p> <p>(1) Shromáždit vícejazyčný vzorek zlatých dat naskenovaných textů a jejich přepisů. Tento cíl byl splněn bez výhrad. Plnění zahrnovalo rešerši zdrojů, které by bylo možné do vzorku zahrnout, a jejich technických parametrů s ohledem na data samotná a na požadavky zvolených systémů OCR. Studentce se podařilo sestavit pětijazyčnou kolekci beletristických a odborných textů.</p> <p>(2) Vybrat konkrétní systém OCR a analyzovat jeho kvalitu na zlatých datech. Tento cíl byl splněn bez výhrad. Studentka dle jasně specifikovaných kritérií vybrala tři systémy a míry pro jejich hodnocení a provedla evaluační experimenty na slovních jednotkách (tokenech) a znacích.</p> <p>(3) Zapojit procedury počítačového zpracování jazyka do analýzy výstupů OCR. Tento cíl byl splněn s výhradami. Sledování distribuce slovních druhů je jistě správné, ale zapojení nástrojů NLP má mnohem větší potenciál.</p> <p>(4) Uživatelsky přívětivě vizualizovat výstup OCR a jeho analýzu. Tento cíl nebyl splněn. V části Conclusion (str. 41) podává studentka velmi stručné odůvodnění, které se odvolává na chybějící strukturu v použitých datech. Ovšem i data bez struktury lze vizualizovat.</p> <p>Vzhledem k velmi dobrému zpracování částí pokrývajících cíle (1)-(3) doporučuji práci k obhajobě.</p>				

Textová část práce	lepší	OK	horší	nevyhovuje
Formální úprava ... <i>jazyková úroveň, typografická úroveň, citace</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Struktura textu ... <i>kontext, cíle, analýza, návrh, vyhodnocení, úroveň detailu</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Analýza	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vývojová dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Uživatelská dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář Click here to enter text.				

Implementační část práce	lepší	OK	horší	nevyhovuje
Kvalita návrhu ... <i>architektura, struktury a algoritmy, použité technologie</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kvalita zpracování ... <i>jmenné konvence, formátování, komentáře, testování</i>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Komentář Click here to enter text.				

Celkové hodnocení Dobře (spíše lepší)
Práci navrhuji na zvláštní ocenění Ne

Datum 15. srpna 2022

Podpis

