

Posudek bakalářské práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce Valentin Christozov
Název práce Web page analyzer for scraping
Rok odevzdání 2023
Studijní program Informatika
Specializace Programming and software development

Autor posudku RNDr. Kateřina Macková
Pracoviště KTIML

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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Splnění zadání	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input 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"/>
Komentář Každá webová stránka má svou specifickou strukturu a zobrazovaná data se často nenachází přímo ve zdrojovém HTML, ale načítají se dynamicky pomocí různých javascriptových skriptů. Cílem práce bylo vytvořit analyzátor webových stránek, který má za cíl provést analýzu webové stránky a na jejím základě najít nejefektivnější způsob, jak ze stránky automaticky vyscrapovat hledaná data. Tento analyzátor má také za cíl vytvořit přehledné uživatelské prostředí pro různé typy uživatelů, které jim má zjednodušit práci s následným vytvářením web scraperů. Rozsah práce je v textové i implementační části splněný.				

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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Vývojová dokumentace	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Uživatelská dokumentace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Práce je psaná v anglickém jazyce a je rozdělena mimo úvodu a závěru na 3 kapitoly. První kapitola popisuje základní terminologii a vysvětluje jednotlivé pojmy spojené s prací. Druhá kapitola je stěžejní kapitolou celé práce a obsahuje popis použitých technologií a samotný průběh jednotlivých částí analýzy webových stránek, včetně příkladů, ukázek z kódu a popis algoritmů. Poslední kapitola obsahuje uživatelskou dokumentaci a popis, jak analyzátor funguje včetně ukázek využití na Apify platformě. Práce je logicky rozčleněna, text je dobře čitelný a logicky navazující a na slušné jazykové úrovni. Jediné, co bych práci vytkla je, že tam opět chybí důkladnější analýza existujících řešení, postupů, diskuze možných algoritmů a zdůvodnění jejich volby. V úvodu je stručně zmíněno, že práce navazuje na předchozí Web Scraper Analyser napsaný v Apify, ale nikde není uvedeno, zda existují i další řešení a jaká. Programátorská dokumentace se nachází přiložená v kódu, ale její úroveň je slabší. Je velice stručná, lehce nepřehledná a obsahuje drobné překlepy a v samotném kódu občas chybí dokumentační komentáře u funkcí.</p>				

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 type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Stabilita implementace	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Komentář Implementační část práce je dobrá. Program po spuštění načte vstupní webovou stránku a hledaná klíčová slova a na výstupu vrací nalezené výskyty a způsoby, jakým je možné je ze stránky získat. Kvalita návrhu co se týče architektury a objektového návrhu je pěkná a až na drobné chyby, které by se daly odladit důkladnějším a pořádnějším testováním, funguje dobře a spolehlivě. Jedinou výtku bych měla k tomu, že pro zadávání vstupu je potřeba používat platformu Apify, nebo vstup zadávat přímo do kódu ve formátu json. Vhodnější by bylo software mít jako samostatnou aplikaci, která má na frontendu okénko na vstup, kam uživatel může zadávat pouze webovou stránku a klíčová slova a konverzi do formátu json už by nemusel řešit.</p> <p>Kvalita návrhu je slušná, jsou zde použity moderní technologie a algoritmy. Kvalita zpracování je už slabší, kód je sice logicky členěn a formátování a jmenné konvence jsou konzistentní, ale má pár menších nedostatků. V textu práce je psáno, že v kódu jsou jasně označeny části, které byly převzaty z původního scraperu, ale v kódu tyto části označené nejsou. Další výtku bych měla k výše zmíněné absenci dokumentačních komentářů a ke kvalitě programátorské dokumentace. Zároveň se v kódu nacházejí také nevyřešená 'TODO'. Nicméně implementace je stabilní a funguje podle očekávání.</p>				

Celkové hodnocení Dobře (spíše lepší)

Práci navrhuji na zvláštní ocenění Ne

Datum 31. května 2023

Podpis