

Posudek práce

předložené na Přírodovědecké fakultě JU

- posudek vedoucího
 bakalářské práce
- posudek oponenta
 diplomové práce

Autor/ka: **Bc. Radek Hiess**
Název práce: **Obsahová analýza webových stránek a hlasová syntéza**
Studijní program a obor: Aplikovaná informatika
Rok odevzdání: 2020

Jméno a tituly vedoucího/opponenta: Ing. Jiří Jelínek, CSc.

Pracoviště: Ústav aplikované informatiky

Kontaktní e-mail: jjelinek@prf.jcu.cz

Odborná úroveň práce:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Věcné chyby:

- téměř žádné vzhledem k rozsahu přiměřený počet méně podstatné četné závažné

Výsledky:

- originální původní i převzaté netriviální kompilace citované z literatury opsané

Rozsah práce:

- veliký standardní dostatečný nedostatečný

Grafická, jazyková a formální úroveň:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Tiskové chyby:

- téměř žádné vzhledem k rozsahu a tématu přiměřený počet četné

Celková úroveň práce:

- vynikající velmi dobrá průměrná podprůměrná nevyhovující

Slovní vyjádření, komentáře a připomínky vedoucího/oponenta:

Cílem práce bylo provedení rešerše v oblasti podpory osob s poruchou zraku (OsPZ) při procházení webu a vytvoření vlastní metodiky pro zpracování obsahu webových stránek a jeho transformaci do podoby přijatelné pro tyto osoby. Dílčím cílem bylo využití již existujících služeb pro tuto činnost. Součástí práce byla i realizace pilotní aplikace implementující navrženou metodiku a analýza získaných výsledků.

Práce je rozvržena do 6 kapitol. V úvodu jsou představeny cíle práce a motivace k volbě tohoto tématu. První kapitola obsahuje avizovanou rešerši SW pro podporu OsPZ. Zkoumáno je několik kategorií těchto produktů. Druhá kapitola popisuje existující služby pro analýzu textu a obrazu a syntézu řeči. Její součástí jsou i popisy metod, které se pro zpracování těchto vstupů používají. Zkoumanými službami byly produkty firem Microsoft (Microsoft Cognitive Services), IBM (Watson), Google (Google Cloud Platform) a Amazon (Amazon Web Services).

Třetí kapitola obsahuje návrh vlastní metodiky zpracování obsahu webové stránky za použití vybraných výše analyzovaných služeb. Jádrem metodiky je pak integrace jednotlivých získaných výstupů do finálního celku. Po průzkumu možností byla pro sumarizaci textu použita i velmi aktuální a pokročilá technika BERT využívající neuronové sítě typu Transformer. V rámci předzpracování obsahu jsou využívány i doplňkové informace o struktuře stránky.

Čtvrtá kapitola se zabývá vývojem pilotní aplikace ověřující navrženou metodiku. Jako vývojové prostředí bylo zvoleno Node-RED umožňující vývoj tzv. flow-based síťových aplikací pomocí rychlého prototypování. Definovány jsou i funkční a nefunkční požadavky. GUI aplikace je velmi jednoduché, což odpovídá jejímu pilotnímu určení.

Testování aplikace je popsáno v kapitole 5. Rozsah zejména uživatelského testování je vcelku omezený, počet testerů mohl být vyšší. Výsledky šetření jsou shrnuty v kap. 6. Chyba komunikace s knihovnou bert_summarizer měla být autorem analyzována a odstraněna stejně jako další zmíněné technické problémy. I přesto je výsledné hodnocení metodiky při práci s textem v porovnání s konkurenčními produkty velmi dobré, což dokazují hodnoty v tab. 2. Rezervy jsou zejména v přesnosti popisu. Autor uvádí, že porovnání v oblasti zpracování obrázku nemohlo být provedeno, což rovněž zdůvodňuje. Bylo by však vhodné najít jinou aplikaci pro porovnání a tuto část testování dopracovat. Navržená aplikace splnila svou pilotní roli.

Práce má odpovídající formální, grafickou a textovou kvalitu a obsahuje 34 odkazů na relevantní převážně zahraniční zdroje. V elektronické příloze práce jsou k dispozici zdrojové kódy a podpůrné knihovny.

S výše uvedenými připomínkami doporučuji práci k obhajobě a navrhuji hodnocení výborně nebo velmi dobře podle výsledků obhajoby.

Případné otázky při obhajobě a náměty do diskuze:

1. Předved'te systém v činnosti.

Práci

doporučuji

nedoporučuji

uznat jako diplomovou/bakalářskou.

Navrhuji hodnocení stupněm:

výborně velmi dobře dobře neprospěl/a

Místo, datum a podpis vedoucího/oponenta:

V Českých Budějovicích dne 29. června 2020

.....