

Posudek práce

předložené na Přírodovědecké fakultě JU

- posudek vedoucího
 bakalářské práce
- posudek oponenta
 diplomové práce

Autor/ka: **Patrik Dohnal**
Název práce: **Detekce kategorie obsahu webové stránky prostřednictvím metod strojového učení**
Studijní program a obor: Aplikovaná informatika
Rok odevzdání: 2021

Jméno a tituly vedoucího/opponenta: **Ing. Jan Fesl, Ph.D.**

Pracoviště: Katedra informatiky

Kontaktní e-mail: jfesl@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:

Předložená bakalářská práce obsahuje dobře zdokumentovaný proces aktuálních možností pro automatickou klasifikaci obsahu webové stránky (AKOWS). AKOWS je problematika, kterou neustále řeší přední technologické firmy.

Z hlediska struktury – tj. po textové a grafické stránce je práce kvalitní a nemám k ní zásadní výhrady.

Po obsahové stránce je práce celkově homogenní, jednotlivé pasáže se navzájem dobře doplňují. Student ve své práci shrnul dosavadní známé metody, které popsal matematickou i triviální formou. Výčet metod je kompletní až na pokročilé deep-learningové metody zasahující do oblasti NLP, které nebylo kvůli časovému presu možné do práce zintegrovat.

Student věnoval značné úsilí pro vytváření datové sady, která nebyla jednoduše získatelná a nad touto implementoval a otestoval účinnost 3 základních klasifikačních strategií. Pro navržený model dosáhl velmi slušné úspěšnosti.

Při hodnocení práce jsem zohlednil opravdu velkou časovou náročnost vytváření kvalitní datové sady, která studentovi zabránila ve vyzkoušení pokročilejších metod založených např. na modelech využívající reprezentace prostřednictvím modelů vycházející z technik word2vec či doc2vec.

Pro studenta vidím jako perspektivní v práci pokračovat i na úrovni magisterského studia.

Případné otázky při obhajobě a náměty do diskuze:

- 1) V kterých Vámi nepoužitých metodách strojového učení spatřujete možnost vylepšení Vámi dosažených výsledků?
- 2) Kolik času Vám zabralo vytvoření datové množiny?

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 17. května 2021