



Posudek oponenta závěrečné práce

Oponent práce: Ing. Karel Klouda, Ph.D.
Student: Bc. Martin Macho
Název práce: Využití strojového učení a AI pro extrakci informací z webu
Obor / specializace: Znalostní inženýrství
Vytvořeno dne: 26. srpna 2021

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Všechny body zadání byly splněny a to velmi poctivě. Zadání vyžadovalo využití několika různých technologií a jeho největší obtížnost a pracnost spočívala hlavně v získání z přípravě dat.

2. Písemná část práce

95 / 100 (A)

Text práce, který je psán v češtině, je velmi přehledně členěn a velmi dobře srozumitelný. Popis nástroje pro získávání obrázků webových stránek se četl velmi dobře. Čtenář byl seznámen nejen s problémem samotným, ale i s různými možnostmi jak jej řešit a obtížemi, které jednotlivé cesty řeší či neřeší. Velmi rozumně je pojatá i rešerše modelů pro detekci objektů v obrázcích, takže čtenář následně může celkem snadno porozumět dalším kapitolám. Kapitola u vyhodnocení experimentů by mohla být místy trochu návodnější, ale i tak je dobře srozumitelná.

Pár drobných nedostatků jsem našel (např. vzorec (3.2) na str. 23, nevysvětlená proměnná p ve vzorci (3.4) na str. 29,), ale celkově obsahuje práci mizivě množství jakýchkoli prohřešků.

3. Nepísemná část, přílohy

90 / 100 (A)

Jelikož práce kombinovala využití několika nástrojů a předtrénovaných modelů, jsou v nepísemné části většinou zdrojové kódy krátkých skriptů v jazyce Python či JavaScript. Důležitým výstupem práce je též dataset, který obsahuje obrázky stránek spolu s popisnými oblastmi a HTML zdrojovými kódy.

4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Jedná se o zajímavý experiment, který ukazuje, jak asi může být úspěšná snaha o získávání strukturovaných dat z obrázku webové stránky a v neposlední řadě ukazuje i způsob, jak takový obrázek získat. Experimenty s modely by mohly být obsáhlejší, ale nedomnívám se, že by další experimenty mohly vést k výrazně lepším výsledkům.

Celkové hodnocení

90 /100 (A)

Jedná se o pěknou práci obsahující zajímavý experiment, který z pohledu využití neuronových sítí dává velmi dobrý smysl. Ukazuje se, že využití není bezproblémové, ale v některých situacích (nejsou k dispozici strukturovaná data přímo od autorů stránek) může dávat smysl.

Otázky k obhajobě

1. Neexistují nějaké experimenty, které by měli stejný cíl jako Vaše práce, ale využívali by pouze textová data (text mining)? Např. detekce toho, jestli je produkt skladem, by mohla být úspěšná.
2. Jak přesně byl vytvořen náhodný seznam e-shopů, o kterém se píše na str. 48?

Instrukce

Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.