FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	prof. Ing. RNDr. Martin Holeňa, CSc.
Student:	Bc. Matyáš Skalický
Název práce:	Párování relativního rozvržení pro extrakci dat z dokumentů
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	2. června 2022

Hodnotící kritéria

1. Splnění zadání

▸ [1] **zadání splněno**
[2] zadání splněno s menšími výhradami
[3] zadání splněno s většími výhradami
[4] zadání nesplněno

Zadání pokládám za splněné.

2. Písemná část práce 70 /100 (C)

Písemná část práce má značný počet nedostatků ve všech směrech: pokud jde o logickou strukturu stejně jako i pokud jde o obsah, věcnou správnost, citování zdrojů a srozumitelnost pro čtenáře. Konkrétně:
1. Sekce Implementaion details by logicky patřila před sekci Experiments, ne až za ni.
2. Od sekce Related Work se očekává, že podá přehled literatury týkající se řešeného problému, ne že bude vysvětlovat základy metod použitých při jeho řešení, k tomu by práce měla mít samostatnou sekci, např. Methodology nebo Background (do sekce Background by zapadal i obsah sekce o dokumentech umístěné před Related Work).
3. Cíle práce nejsou explicitně zformulovány.
4. K výběru z vice možností bylo místo křížové validace použito pouze jednoduché rozdělení na trénovací a testovací data.
5. U klíčové reference [12] chybí údaj o tom, na kterou konferenci byla zaslána (je sice uvedený v textu práce, ale obsahově patří do seznamu literatury).
6. Přestože se v práci používá uspořádání citací, odkazuje se na ně někdy i neuspořádaně, např. [2,29,3,4].
7. Seznam zkratek neobsahuje všechny zkratky vyskytující se v práci, nenašel jsem v něm zkratku IOU.

3. Nepísemná část, přílohy 90 /100 (A)

Diplomant provedl kvalitní výzkum, je pouze škoda, že k němu používal pouze proprietální dataset, neboť to snižuje význam experimentální části z hlediska diplomky (byť nikoliv z hledika Rossumu).

4. Hodnocení výsledků, jejich využitelnost 95 /100 (A)

Využitelnost výsledků je na diplomovou práci výrazně nadprůměrná. O tom svědčí už to, že z ní vznikl článek na konferenci, což je i u výzkumných diplomek menšinový jev. Navíc se domnívám, že díky přímé souvislosti tématu práce se zaměřením Rossumu jsou její výsledky využitelné i tam, včetně těch negativních.

Celkové hodnocení 85 /100 (B)

Přestože písemná část práce má mnoho nedostatků ve všech směrech, Matyáš Skalický svou diplomkou prokázal, že má schopnosti pro nadprůměrně kvalitní výzkum. Byla by škoda, pokud by nepokračoval do doktorského studia.

Otázky k obhajobě

K práci jsem měl řadu otázek, ale diplomant byl tak hodný, že mi je přišel všechny osobně zodpovědět.

Instrukce

Splnění zadání

Posuďte, 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, posuďte 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 posuďte, 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. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posuďte, 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.