

Posudek oponenta diplomové práce

Student: Jarolím Jordán, Bc.

Téma: Analýza a získávání informací ze souboru dokumentů spojených do jednoho celku (id 20912)

Oponent: Bartík Vladimír, Ing., Ph.D., UIFS FIT VUT

- 1. Náročnost zadání** **obtížnější zadání**
Zadání vyžadovalo nastudování většího množství různých technologií souvisejících se zpracováním dokumentů, např. technologie OCR nebo hodnocení textové i vizuální podobnosti dokumentů. Cílem poté bylo využití těchto technologií v aplikaci, která umožní na základě obsahu jednotlivých stránek rozdělení souborů obsahujících více naskenovaných dokumentů na jednotlivé dokumenty.
- 2. Splnění požadavků zadání** **zadání splněno**
Všechny body zadání byly splněny.
- 3. Rozsah technické zprávy** **splňuje pouze minimální požadavky**
Rozsah technické zprávy se pohybuje na dolní hranici povoleného rozsahu. Některé části (viz níže) mohly být rozepsány více, nic zcela zásadního však nechybí.
- 4. Prezentací úroveň předložené práce** **70 b. (C)**
Návaznost jednotlivých kapitol je správná, rozsahem však teoretická část výrazně převyšuje část praktickou. Teoretická část je velmi kvalitní a nemám k ní žádné výhrady. Myslím si však, že dále o návrhu aplikace, volbě vývojového prostředí a samotné implementaci mohlo být napsáno více. Z práce také není úplně jasný způsob vytvoření rozhodovacího stromu na str. 34. Naopak teoretická část je napsána srozumitelně a poměrně podrobně. V kap. 5.4. postrádám podrobnější popis datasetu, na kterém byla úspěšnost výsledné aplikace ověřována.
- 5. Formální úprava technické zprávy** **82 b. (B)**
Po formální stránce nejsou nedostatky nijak velké. Chyb a překlepů není příliš mnoho, vytknout by se ještě dalo místy zbytečné používání anglických pojmů i v případech, kdy existuje český ekvivalent (např. document clustering - shlukování dokumentů).
- 6. Práce s literaturou** **90 b. (A)**
Po stránce práce s literaturou nemám k práci výhrady, vše potřebné je správně a v souladu s normami citováno.
- 7. Realizační výstup** **90 b. (A)**
Realizační výstup představuje kvalitní aplikaci, která umožňuje rozdělit PDF soubor, vzniklý např. hromadným skenováním na jednotlivé dokumenty, které jsou v tomto souboru obsaženy. Úspěšnost je velmi dobrá, aplikace i tak umožňuje ručně opravit případné chybné rozdělení dokumentu. Výsledky experimentů jsou v práci také popsány, i když tato část mohla také být obsáhlejší.
- 8. Využitelnost výsledků**
Největším přínosem této diplomové práce je vytvořená aplikace. Známa aplikace s podobnou funkcí dosud neexistuje.
- 9. Otázky k obhajobě**
 - Uvažoval jste o využití tzv. stemmingu při předzpracování textu? Mělo by to pro vaši úlohu význam?
 - Popište, jakým způsobem byl vytvořen rozhodovací strom na str. 34.
- 10. Souhrnné hodnocení** **82 b. velmi dobře (B)**
Průměrná technická zpráva je vyvážená velmi kvalitním realizačním výstupem, který je přínosem. Navrhuji proto hodnotit tuto diplomovou práci stupněm B (velmi dobře).

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 7. června 2018

.....
podpis