

Posudek oponenta diplomové práce

Student: Ondřej Karel, Bc.
Téma: Klasifikace vztahů mezi pojmenovanými entitami v textu (id 22682)
Oponent: Doležal Jan, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání téměř splněno s drobnými výhradami
Zadání je splněno s drobnými výhradami.

Student v technické zprávě nepopisuje způsob využití dodatečných informací ke zvyšování přesnosti což je část bodu dva. Píše pouze o vstupním formátu MG4J, který dodatečné informace obsahuje.

Dále jsem v technické zprávě nenašel jakým způsobem bylo implementováno pravidelné rozšiřování extrahovaných dat na základě nově získávaných webových dat. Student pouze uvádí, že se opírá o systém, jenž vytvořil jako svou bakalářskou práci. V kapitole "Návrh a implementace" pouze píše, že se od něj nechal inspirovat v návrhu konfiguračního souboru.
3. **Rozsah technické zprávy** splňuje pouze minimální požadavky
Rozsah technické zprávy splňuje pouze minimální požadavky. Všechny části technické zprávy jsou ale informačně bohaté a pro práci skutečně nezbytné. V kapitole "Návrh a implementace" postrádám některé informace ohledně návrhu a implementaci. Např. mi schází odůvodnění, proč si zvolil nástroj NLTK pro extrakci vět z čistého textu.
4. **Prezentační úroveň předložené práce** 80 b. (B)
Logická struktura technické zprávy je dobrá. Práce je čtivá, dobře pochopitelná pro čtenáře a jednotlivé kapitoly na sebe dobře navazují. Kresby vytvořené studentem dále vysvětlují popsané skutečnosti.

Mám ale výhrady k tomu, že student sloučil návrh a implementaci do kapitoly "Návrh a implementace" kde si nejsem tak jist, zda-li píše ohledně návrhu či implementaci.
5. **Formální úprava technické zprávy** 95 b. (A)
Formální úprava technické zprávy je dobrá. Nalezl jsem pouze jednotky překlepů. Kresby vytvořené studentem jsou ve vektorovém formátu, písmem zapadají do zbytku práce.
6. **Práce s literaturou** 100 b. (A)
Výběr studijních pramenů je výborný a student se na ně v celé práci vhodně odkazuje. Z technické zprávy je patrné, že si student problematiku detailně nastudoval.
7. **Realizační výstup** 99 b. (A)
Zdrojové kódy realizačního výstupu jsou dostatečně okomentovány. Student vytvořil svou práci modulární a snadno rozšiřitelnou, což uvádí v technické zprávě. Využívá validaci konfiguračního souboru což zabraňuje zbytečným chybám při tvorbě rozšíření. Výtku mám k souboru README, jenž obsahuje chybu příkladu spuštění systému (domnívám se, že správná cesta je "src/re").
8. **Využitelnost výsledků**
Jedná se o práci kompilačního charakteru. Práce je využitelná pro další výzkum a experimenty v rámci výzkumné skupiny KNOT.
9. **Otázky k obhajobě**
 - Jakým způsobem byly využity dodatečné informace ke zvyšování přesnosti? (část bodu 2 ze zadání)
 - Jakým způsobem bylo implementováno pravidelné rozšiřování extrahovaných dat na základě nově získávaných webových dat? (bod 3 ze zadání)
10. **Souhrnné hodnocení** 89 b. velmi dobře (B)
Student odvedl velmi pěknou práci a opírá se o aktuální literaturu. Technická zpráva ale obsahuje nedostatky, které vyžadují vysvětlení (viz bod 9. Otázky k obhajobě). Kapitulu "Návrh a implementace" by bylo vhodné rozdělit a dále rozvést. Z těchto důvodů navrhuji hodnocení stupněm B (89 bodů).

V Brně dne: 29. června 2020

Doležal Jan, Ing.
oponent