

Posudek oponenta bakalářské práce

Student: Križan Samuel
Téma: Zjednodušování pojmenovaných entit ve slovenštině (id 24563)
Oponent: Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno
3. **Rozsah technické zprávy** splňuje pouze minimální požadavky
Rozsah technické zprávy splňuje pouze minimální požadavky.
4. **Prezentační úroveň předložené práce** 60 b. (D)
Prezentační úroveň práce je spíše špatná. Technická zpráva je místy hůře pochopitelné pro čtenáře. Některé její části jsou vysvětleny velmi stroze (např. podkapitola 3.6) a bylo by vhodné je rozšířit. Velkým problémem je, že student v práci neuvádí příklady, které by pomohly k lepšímu pochopení. Práce obsahuje několik zavádějících tvrzení a některé pojmy zůstávají nevysvětleny (např. povrchová slova, rozlišovací stránky atd.). Některé tabulky jsou nepřehledné (např. tabulka 5.2, 5.3 a 5.4). Kapitola 5 - Návrh a implementácia systému se věnuje spíše popisu již existujícího systému pro češtinu, než návrhu vlastního systému. Není zde ani uvedeno, v jakém programovacím jazyce bude práce vytvořena.
5. **Formální úprava technické zprávy** 75 b. (C)
Formální úprava technické zprávy je průměrná. Jazykovou stránku technické zprávy nemohu hodnotit, protože nejsem rodilý mluvčí. Po typografické stránce by práci prospělo, kdyby některé její části byly lépe formátovány (např. hypertextové odkazy nebo některé tabulky).
6. **Práce s literaturou** 85 b. (B)
Výběr studijních pramenů je velmi dobrý. Student v práci čerpá z velkého množství vědeckých publikací, které v práci vhodně cituje. Problém je, že student většinu nabytých poznatků v práci nijak nevyužil.
7. **Realizační výstup** 55 b. (E)
Student v rámci bakalářské práce vytvořil systém pro extrakci entit ze slovenské Wikipedie a systém pro automatické doplňování tvarů jmen. Při implementaci student vycházel do značné míry z již existujícího systému pro český jazyk. Na straně 15 technické zprávy student uvádí, které moduly vytvářel sám a které ne. Tento obrázek je však nepřesný, protože např. "Modul pre tvorbu znalostnej bázy" nevytvářel zcela sám, ale upravoval již existující modul pro český jazyk. Studentem vytvořené moduly jsou komentovány velmi stroze. K implementovanému systému chybí jakákoliv dokumentace. Vyhodnocení systému je velmi povrchní.
8. **Využitelnost výsledků**
Jedná se o práci implementačního charakteru. Výsledky práce studenta mohou být pravděpodobně použity v rámci výzkumné skupiny KNOT@FIT.
9. **Otázky k obhajobě**
 - Na straně 20 píšete, že při tvorbě znalostní báze dochází k přetěžování API Wikipedie. Proč tedy používáte API a nezískáváte potřebné informace z offline souborů, které by pro Vaše potřeby měly být dostupné?
 - Na straně 25 uvádíte, že při aktualizaci znalostní báze přidáváte unikátní entity ze staré verze Wikipedie do nové verze. Nemůže mít jejich odstranění racionální důvod (např. porušení pravidel Wikipedie)? Můžete porotě ukázat několik příkladů odstranění entit z Wikipedie a uvést pravděpodobné důvody, proč k tomu došlo?
10. **Souhrnné hodnocení** 60 b. uspokojivě (D)
Zadání práce bylo splněno v minimální možné míře. Student se příliš nezabýval návrhem vlastního systému a omezil se většinou na úpravu již existujícího systému pro český jazyk. Studentem předložený systém bude pravděpodobně možné použít v projektech výzkumné skupiny KNOT@FIT. Rozsah technické zprávy splňuje pouze minimální požadavky a je zde prostor k rozšíření. Prezentační úroveň práce je spíše špatná. Zcela chybí dokumentace k systému. Z těchto důvodů navrhuji hodnocení stupněm D.

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

V Brně dne: 2. června 2022

Otrusina Lubomír, Ing.
oponent