

## Posudek oponenta bakalářské práce

**Student:** Dočekal Martin  
**Téma:** Pokročilé metody strojového učení pro klasifikaci textu (id 19757)  
**Oponent:** Otrusina Lubomír, Ing., UPGM FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
2. **Splnění požadavků zadání** zadání splněno  
Zadání práce bylo splněno. Student se seznámil s dostupnými metodami a nástroji a na základě získaných znalostí navrhl a realizoval systém pro sémantickou klasifikaci obsahu knih a dalších zadaných dokumentů.
3. **Rozsah technické zprávy** je v obvyklém rozmezí  
Rozsah technické zprávy je v obvyklém rozmezí. Všechny části technické zprávy jsou informačně bohaté a pro práci skutečně nezbytné.
4. **Prezentační úroveň předložené práce** 65 b. (D)  
Prezentační úroveň technické zprávy je dobrá. Práce je pochopitelná pro čtenáře, ale je zde několik nedostatků. Některé pasáže (např. učení s/bez učitele, křížová validace) jsou v textu vysvětleny nedostatečně. Některé pojmy jsou vysvětleny méně srozumitelně nebo jsou špatně začleněny do textu. U většiny algoritmů chybí popis jejich vlastností. Tabulka 2.1 není v práci vysvětlena vůbec.
5. **Formální úprava technické zprávy** 75 b. (C)  
Formální úprava technické zprávy je dobrá. Vytkl bych snad jen občasné překlepy, přítomnost neexistujících slov (např. oindexovali), případně méně srozumitelné obrázky/diagramy, které nejsou příliš standardní. Obrázky jsou často také obtékány textem, což ubírá práci na přehlednosti.
6. **Práce s literaturou** 75 b. (C)  
Student použil velké množství zdrojů, které jsou relevantní k danému tématu. Často se jedná o stránky různých technologií, zhruba polovinu tvoří seriózní vědecké publikace. V některých případech ale není zcela jasné, odkud student informace čerpal (např. některé sekce kapitoly 2).
7. **Realizační výstup** 85 b. (B)  
Realizační výstup je na velmi dobré úrovni. Student implementoval systém, který dobře klasifikuje české dokumenty. Vytkl bych snad jen fakt, že student při vyhodnocení srovnává svůj systém s nástrojem fastText, který není v práci nikde popsán.
8. **Využitelnost výsledků**  
Jedná se o práci kompilačního charakteru. Realizační výstup práce je aktuálně využíván v projektu NAKI CPK, který se zabývá využitím sémantických technologií pro zpřístupnění kulturního dědictví prostřednictvím Centrálního portálu knihoven. Tato skutečnost ale není bohužel v práci dostatečně zmíněna.
9. **Otázky k obhajobě**
  - V práci uvádíte, že výstupy z nástroje Doc2Vec byly klasifikovány pomocí k nejbližších sousedů, pro k rovno 3. Můžete komisi zdůvodnit, proč jste vybral zrovna tuto hodnotu? V práci to nikde uvedeno není.
  - V kapitole 5 uvádíte, že nejhorších výsledků bylo dosaženo za použití nástroje FastText. Máte představu, proč tento nástroj dosahoval tak špatných výsledků v porovnání s dalšími algoritmy?
10. **Souhrnné hodnocení** 75 b. dobře (C)  
Jednalo se o průměrně obtížné zadání. Kvalita technické zprávy je dobrá, ale trpí některými nedostatky. Naopak kvalita realizačního výstupu je velmi dobrá. Realizační výstup je používán v rámci projektu NAKI CPK. Z těchto důvodů hodnotím práci stupněm C.

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

V Brně dne: 1. června 2017

.....  
podpis

