



# Posudek oponenta závěrečné práce

Oponent práce: Ing. Magda Friedjungová, Ph.D.  
Student: Bc. Martin Bendík  
Název práce: Automatická detekce témat v básnických textech  
Obor / specializace: Znalostní inženýrství  
Vytvořeno dne: 4. června 2023

## Hodnotící kritéria

### 1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání práce bylo splněno.

### 2. Písemná část práce

83 /100 (B)

Práce je psána v angličtině a je logicky členěna. Jednotlivé kapitoly jsou informačně vyvážené. Teoretická část je poměrně rozsáhlá, první čtyři kapitoly popisují jak nesupervizované, tak i semi a supervizované metody pro oblast NLP a modelování/klasifikaci témat. Zbývající kapitoly jsou věnovány praktické části. Písemná podoba práce (včetně vizualizací) se mi líbí, zmínila bych však několik drobností.

- Práce obsahuje minimum překlepů (např. "acuracy" u Fig. 6.11 a 6.12).
- Hodilo by se přidat prolinkování na jednotlivé kapitoly v "Structure of the thesis".
- Student předkládá seznámení se se základní teorií v kapitole jedna a související práce. Uvítala bych zde více referencí na odbornou literaturu, první se vyskytuje až v druhé části 1.3 Related work. Resp. i v dalších částech bych uvítala reference, pokud jsou zmíněny nějaké pojmy nebo výroky, např. při první zmínce metody BERT (str. 9). Není to ale tak, že by reference chyběly, jen se vyskytují méně a později, než bych doporučila. V některých případech je referenci vhodné uvést přímo za "pojmem", např. Röder et al. [9], než na konci věty/odstavce (str. 10).
- V sekci 2.1 Evaluation by se hodily i vzorečky pro zmíněné metriky.
- V sekci 2.2 by se hodilo ukázat i histogramy, aby byla prezentace metod konzistentní.
- Na obrázku 6.5 je 6 embedding modelů, v textu jsou popsány pouze některé - nebála bych se popsat všechny použité modely.
- U většiny grafů je podseknutá osa y.

- V textu nejsou zavedeny zkratky, např. NB (str. 54), SVC (str. 55), nicméně jsou vysvětleny v příloze A.

### 3. Nepísemná část, přílohy

93 /100 (A)

Student zvolil přiměřené technologie (Python, Jupyter Notebook) a knihovny. Kód je přehledný a dostatečně komentovaný. Také obsahuje (zdá se) podklad pro všechny vizualizace, které byly v písemné práci prezentovány. Experimenty jsou opakovatelné a kód lze využít pro další práci a nasazení "do praxe". Snad jen - v "readme" student uvádí, že prefix v názvu jednotlivých notebooků značí chronologické pořadí pro spuštění, ale některá čísla chybí (stejně tak u skriptů).

### 4. Hodnocení výsledků, jejich využitelnost

90 /100 (A)

Student v rámci práce předložil několik modelů jak pro nesupervizovanou, tak pro (semi-)supervizovanou úlohu modelování témat, spolu s vyhodnocením a vizualizací výsledků. Předložené výsledky u (semi-)supervizované úlohy jsou na použitých datech první tohoto druhu. Výsledky přinášejí několik zajímavých poznatků, které student vhodně prezentuje a komentuje (včetně zvolených metod, např. dobrá argumentace pro doc2vec embedding model). Student ve své práci také navázal na výsledky BP Anny Tesaříkové, která se zabývala nesupervizovanou úlohou. Na práci lze navázat dalšími experimenty (student uvádí několik případů, kterým by bylo možné se dále věnovat), případně lze uvažovat i o jejich publikování. Praktické využití práce zřejmě najde v rámci řešeného projektu s ÚČL.

### Celkové hodnocení

92 /100 (A)

Předloženou práci navrhuji klasifikovat stupněm A a to z výše zmíněných důvodů.

### Otázky k obhajobě

Na straně 37 uvádíte: "Later, just before submitting the thesis, there was more data annotated by the experts, but we did not have time to run and evaluate the experiments again. However, the created source codes allow to run experiments with new data and evaluate the results." Spustil jste po odevzdání práce experimenty s novými daty? Pokud ano, jakých jste dosáhl výsledků?

## **Instrukce**

### **Splnění zadání**

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

Posudte, 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.