



# Hodnocení vedoucího závěrečné práce

**Vedoucí práce:** Ing. Magda Friedjungová, Ph.D.  
**Student:** Jonáš Sirko  
**Název práce:** Vizualizace extrakce témat Korpusu českého verše  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 7. července 2023

## Hodnotící kritéria

### 1. Splnění zadání

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

### 2. Písemná část práce

65 /100 (D)

Práce je víceméně logicky členěna, nicméně stručnějšiho rázu - dovedu si představit obsáhlejší jak teoretickou, tak praktickou část. Student stručně představuje několik metod pro modelování témat jako je LDA, top2vec a některé z částí BERTopicu (např. BERT nebo UMAP nejsou vysvětleny), dále pak stručně představuje základní shlukovací metody a s nimi dvě související vizualizační metody. Teorie je místy poněkud stručná a jednotlivé sekce by se daly rozšířit i o další metody, především co se týče vizualizací. Místy chybí zadefinování pojmů (např. bag of words, token v obr. 2.1 apod.). V praktické části pak student využívá např. nástroj LDAvis, který by se hodilo více popsat. Stejně tak student pracuje s metodami pro redukci dimenzionality, ač to lze považovat nad rámec zadání, uvítala bych stručné představení použitých metod (např. t-SNE, PaCMAP). Dále by se hodilo zavést metriky, kterými byly metody vyhodnocovány (např. silhouette score). V teoretické části byla představena matice vzájemných vzdáleností, kterou student v praktické části nepoužívá. Naopak použitý wordcloud není v teoretické části popsán. Část s experimenty a diskuzí je dobrá, student v závěru představuje barevné zvýraznění top n shluků a možnou vizuální prezentaci výsledků z různých modelů.

Student v závěru práce konstatuje, že určité nedostatky práce spočívají v menším množství zkoumaných shlukovacích metod. Nicméně by bylo vhodné se i více pověnovat vizualizacím, na které zadání práce cílilo.

Typografická stránka práce by mohla být lepší (matematické zápisy (např. záměna symbolu pro násobení se symbolem pro konvoluci), formátování), stejně tak práce

obsahuje několik překlepů a gramatických chyb. Místy jsou použity nezavedené zkratky, např. RSS, seznam zkratk je nekompletní. Základní statistiky v obrázku 2.5b neobsahují počet nezařazených básní jako je tomu u 2.6b a 2.7b. Pro lepší srovnání by se hodilo mít statistiky v jedné společné tabulce. Apod.

### 3. Nepísemná část, přílohy

80 /100 (B)

Praktická část se skládá z několika jupyter notebooků a jiných souborů. Jupyter notebooky mají samovysvětlující názvy, ale hodilo by se readme s popisem ostatních souborů. Jupyter notebooky nejsou samy o sobě spustitelné, je potřeba menších úprav. Kód není nijak komentovaný, ale je lehce interpretovatelný. Dovedu si představit, že nějaké části využijeme v rámci projektu TAČR, který se modelování témat v poezii věnuje.

### 4. Hodnocení výsledků, jejich využitelnost

70 /100 (C)

Práce se věnuje modelování témat, které student doplňuje několika vizualizacemi. Výsledky modelování jsou samy o sobě dobré, celý proces by však mohl být doprovázen více vizualizacemi. Nicméně i poskytnuté vizualizace nejspíš půjdou nějakým způsobem znovu použít v řešeném TAČR projektu, respektive by se dalo na studentův kód navázat a doladit navržené řešení (především závěrečnou prezentaci výsledků shlukování).

### 5. Aktivita studenta

- [1] výborná aktivita
- [2] velmi dobrá aktivita
- ▶ [3] průměrná aktivita
- [4] slabší, ale ještě dostatečná aktivita
- [5] nedostatečná aktivita

Myslím, že studentovi se zadané téma líbilo a řešení práce ho i bavilo, avšak jeho časové možnosti/odhady nebyly kompatibilní s našimi představami. Student práci párkrát konzultoval a vždy byl připraven. Nicméně největší část práce odvedl před deadline, kdy už nebylo možné dodat zpětnou vazbu.

### 6. Samostatnost studenta

- ▶ [1] výborná samostatnost
- [2] velmi dobrá samostatnost
- [3] průměrná samostatnost
- [4] slabší, ale ještě dostatečná samostatnost
- [5] nedostatečná samostatnost

Student pracoval samostatně.

## Celkové hodnocení

71 /100 (C)

Odevzdanou práci navrhuji hodnotit klasifikačním stupněm C a to z výše zmíněných důvodů. Student odvedl dobrou práci, kdy se ve většině času zcela sám vypořádával se zadanou úlohou.

## **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.

### **Aktivita studenta**

V souvislosti s průběhem a výsledkem práce posudte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

### **Samostatnost studenta**

V souvislosti s průběhem a výsledkem práce posudte schopnost studenta samostatně tvůrčí práce.

### **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.