



# Posudek oponenta závěrečné práce

**Oponent práce:** Ing. Karel Klouda, Ph.D.  
**Student:** Anna Tesaříková  
**Název práce:** Modelování témat české poezie  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 7. června 2022

## Hodnotící kritéria

### 1. Splnění zadání

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

Všechny body zadání byly splněny. Zadání vyžadovalo seznámit se s netriviálními metodami pro topic modelling a jejich aplikaci na dataset s básnickými texty.

### 2. Písemná část práce

85 / 100 (B)

Text je psán velmi dobrou angličtinou a je srozumitelný. Struktura práce je též rozumně zvolena. Používané metody patří většinou ke složitějším metodám strojového učení a tak samozřejmě nebylo možné je bezezbytku popsat. Naopak míra detailu, s jakou jsou pospány, mi přijde adekvátní. Texty některých částí jsou (přiznanými) citacemi z literatury. To občas způsobuje, že obsahují pojmy, které jsou vysvětleny až později v textu (např. LDA na straně 4) nebo se používá jiné značení pro stejné pojmy (např. term-by-document matice  $X$  a  $A$ ). Mnoho pojmů a metod není vysvětleno vůbec, pouze je uveden odkaz na literaturu. To je v mnoha případech v pořádku, ale např. u "coherence score" mi to nepřijde vhodné. U některých odstavců a částí není úplně zřejmé, proč jsou v práci zmíněny. To platí zejm. pro část 1.2.1, která je navíc (opět přiznanou) citací z wikipedie. To, co se v ní píše, není k tématu příliš relevantní. Typografických a formálních chyb je v práci naprosté minimum.

### 3. Nepísemná část, přílohy

90 / 100 (A)

Příloha obsahuje sadu Jupyter notebooků s popsány experimenty. Jupyter notebooky jsou celkem přehledné a není složité experimenty reprodukovat. Využívá se modelů s existujících knihoven, takže kód není nikterak dlouhý a komplikovaný.

#### 4. Hodnocení výsledků, jejich využitelnost

92 /100 (A)

Jelikož se jedná o nesupervizované metody, je těžké vyhodnotit kvalitu výsledků. Práce ale rozhodně poslouží komukoli, kdo by se chtěl o topic modellingu dozvědět více. K tomu jsou nápomocny i ukázky kódu s využitím těchto metod. Z tohoto pohledu je práce rozhodně hodnotná.

#### Celkové hodnocení

90 /100 (A)

Práce má svoje rezervy, ale vzhledem k obtížnosti používaných metod ji navrhuji hodnotit známkou A. Autorka zjevně pročetla mnoho komplikované literatury a ze získaných informací zkomponovala kvalitní a srozumitelný text doprovazený ukázkami použití těchto metod.

#### Otázky k obhajobě

- 1) Mohla byste vysvětlit, jak se počítá a co zhruba měří coherence score?
- 2) Dal by se některý z modelů použít i v situaci, kdy máme např. několik set básní přiřazených k nějakému tématu ručně a chceme najít další básně věnující se tomuto tématu?

## **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.