



# Posudek oponenta závěrečné práce

**Oponent práce:** Ing. Magda Friedjungová, Ph.D.  
**Student:** Ondřej Černý  
**Název práce:** Rozpoznávání pojmenovaných entit v básnických textech  
**Obor / specializace:** Znalostní inženýrství  
**Vytvořeno dne:** 11. 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, práce obsahuje jen několik překlepů a typografických prohrěšků. Např. systém zavádění zkratk není jednotný (někdy v textu, někdy pouze v seznamu zkratk, některé víckrát). Práce obsahuje dostatek referencí, nicméně bych referenci očekávala i tehdy, pokud student zmiňuje poprvé nějakou problematiku nebo model (např. str. 4 LSTM, BERT apod.). Stejně tak je vhodné dodržet jednotný formát odkazování se na literaturu (primárně používá [číslo], ale minimálně dvakrát i (příjmení, rok vydání)[číslo] nebo příjmení (rok vydání)). Další drobnosti zmíním dle jednotlivých kapitol.

Po obsahové stránce nemám žádné zásadní výtky. Student v první kapitole uvádí čtenáře do problematiky Named Entity Recognition (NER). Představení možných řešení v podkapitole 1.2 je poněkud stručné. Některé podkapitoly jsou zbytečně členěny na další sekce, např. 1.4 a 1.4.1 (jedná se vždy pouze o jeden stručný odstavec). V kapitole 2 student představuje použitý dataset, kapitole je stručnější, ale to hodnotím pozitivně. Obrázek 2.1 by mohl mít světlé pozadí (stejně tak i ostatní obrázky s ukázkami kódů). V kapitole 3 již začíná praktická část práce - student popisuje způsob, jakým byly v jeho práci vybírány entity. Navržená pravidla jsou doplněna ukázkami přímo z Korpusu českého verše a zakončena shrnutím o výskytu jednotlivých variant. Kapitola 4 pak pojednává o kategorizaci nalezených entit. V podkapitole 4.1 chybí pro konzistenci popis kategorie "Place", i když je samovysvětlující. Při zmínce Wikipedie/Wikidata by se hodil odkaz v

poznámce pod čarou. Koláčové grafy (např. 4.9 a ostatní) jsou ve 3D hůře čitelné, navíc je tento efekt zbytečný. V kapitole 5 student popisuje zvolený klasifikační model a potřebné teoretické znalosti k jeho pochopení. Popis je stručný, ale věcný. Obrázek 5.8 by bylo lepší v latexu formátovat jako algoritmus. Kapitola 6 popisuje samotnou implementaci včetně trénování modelu a evaluačních metrik. Obrázky 6.3, 6.4 a další mají podseknutou osu y (rozumím motivaci, ale vhodné by bylo využít značení pro zkrácení osy). Matice záměn jsou fajn, možná by mohly být uvedeny i v normalizované podobě. Kapitola 7 shrnuje provedenou práci a navrhuje možnosti, jak na práci dále navázat.

### 3. Nepísemná část, přílohy

92 /100 (A)

Pro svoji praktickou část student zvolil přiměřené technologie a odevzdal několik souborů se strukturovaným a okomentovaným kódem (jak skripty, tak jupyter notebooky), které doplňuje readme. Do readme (obzvlášť u model\_implementation) by se hodilo doplnit pořadí (možná je to pořadí, ve kterém jsou popsány jednotlivé soubory ve složce?), ve kterém je vhodné soubory spouštět pro opakovatelnost experimentů. Součástí nejsou natrénované modely (pravděpodobně kvůli velikosti). Nicméně části, které jsem zkusila, byly funkční.

### 4. Hodnocení výsledků, jejich využitelnost

94 /100 (A)

Výstupem studentovy práce je systém, který je schopný v Korpusu českého verše rozpoznat pojmenované entity s poměrně velkou přesností. Není mi známo, že by již existovala na toto téma nějaká práce nad stejným korpusem, o to zajímavější a přínosnější je studentova aktivita. Student věnoval notné úsilí přípravě dat (především vytváření dat pro supervizované učení) a návrhu pravidel, jak entity identifikovat. Líbí se mi, že výstup studentovy práce je ve stejném formátu (JSON rozšířený o klíč "entity") jako data na vstupu. V závěru implementace student poskytuje ukázkou vizualizace entit v textu, což považuji za velmi užitečný výstup, který by mohl být využitelný v rámci řešeného TAČR projektu. Na práci lze dále navázat, možnosti student shrnuje v závěrečné kapitole.

## Celkové hodnocení

92 /100 (A)

Odevzdanou práci navrhuji hodnotit klasifikačním stupněm A.

## Otázky k obhajobě

K práci nemám žádné otázky.

## **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.