

Posudek vedoucího práce

Název práce: **Tematické modelování publikační činnosti České akademie věd a umění v letech 1890–1910**

Typ práce: diplomová

Řešitel / řešitelka práce: Bc. Filip Kersch

Vedoucí práce: doc. Mgr. Jindřich Marek, Ph.D.

1. Hodnoticí kritéria a bodové ohodnocení

Obsah práce (max. 70 bodů)

Aspekt práce	Komentář	Body
<i>Kvalita shrnutí současného stavu poznání, používání odborné terminologie, zohlednění různých pohledů na téma práce.</i>	Současný stav poznání je popsán odpovídajícím způsobem, užitá terminologie je vhodná.	10
<i>Definice cílů, výzkumná otázka a její zasazení do současného stavu poznání.</i>	Cíle jsou definovány jasně a výzkumná otázka formulována odpovídajícím způsobem ve vztahu k současnému stavu poznání.	10
<i>Metodologické zpracování výzkumné otázky, vysvětlení volby využitých metod s ohledem na výzkumnou otázku.</i>	Metodologické zpracování odpovídá tématu práce a vysvětlení je dostatečné.	20
<i>Analýza, interpretace výsledků a formulace samostatných závěrů vč. zohlednění současného stavu poznání.</i>	Analýza a interpretace výsledků jsou na odpovídající úrovni, podrobnosti v celkovém komentáři k práci.	18
<i>Invence, inovativní a původní přístup.</i>	Student zpracoval téma samostatně a invenčně, dospěl k zajímavým výsledkům.	10
Celkem bodů (obsah)		68

Formální stránka práce (max. 30 bodů)

Aspekt práce	Komentář	Body
<i>Formální struktura práce a návaznost kapitol.</i>	Struktura práce odpovídá tématu a kapitoly na sebe logicky navazují.	10
<i>Aktuálnost zdrojů, korektní citování, cizojazyčné zdroje.</i>	Nemám námitek.	10

<i>Slohové zpracování a gramatická správnost.</i>	Nemám námitek.	10
Celkem bodů (formální str.)		30
CELKEM BODŮ		98

2. Komentář k hodnocení práce

Diplomová práce Filipa Kersche si klade za cíl provést strojovou analýzy publikační činnosti České akademie císaře Františka Josefa pro vědy, slovesnost a umění z let 1890–1910. Pro tyto účely student vyžil postupu, který sám navrhl a který zahrnuje stažení plných textů z digitální knihovny, jejich lemmatizaci, odfiltrování nežádoucích slov (zejm. jedno- a dvouslovných předložek a spojek, zkratků označujících strany apod.) a následnou analýzu na základě tematického modelování.

Práce obsahuje vedle samotné analýzy samozřejmě i úvodní části, které poučeným způsobem představují Českou akademii věd a umění a její publikační činnost, problematiku digitalizace fondů a „makroanalýzy“ digitálních dat. V úvodní části diplomant relevantně vysvětluje, proč ze své analýzy vyloučil monografie podpořené ČAVU. (V Příloze c. 1 potom najdeme seznam publikací vydaných ČAVU v letech 1890-1918 (sic). Záznamy obsahují příjmení (bohužel ne křestní jména) autorů, názvy děl a odkazy na záznam v elektronickém katalogu a digitální knihovnu.)

Analýza témat, která byla v rámci korpusu textů vymezena (celkem jich je 35, přičemž počet je třeba předem stanovit), jednak potvrdila některé předchozí intuitivní závěry ohledně kvantity produkce (i pokud jde o počet témat) v různých oborech (publikační činnost ČAVU zahrnuje celé spektrum tehdejší vědní produkce, přičemž STEM obory měly poměrně bohatou a diverzifikovanou produkci, ale jejich postavení v rámci Akademie bylo přitom podřadné), jednak potvrdila závěry některých dílčích analýz, které se zabývaly některými obory (botanika apod.). Student narazil i na některé kontextuální poznatky, např. že zástupci STEM oborů nechtěli v rámci ČAVU publikovat česky a kvůli dosahu svých univerzálně zaměřených publikací preferovali spíše Královskou českou Společnost nauk, která uveřejňovala i publikace např. v němčině, nebo publikace v zahraničí. V oblasti humanitních věd se projevilo i založení Českého časopisu historického (1895), které vedlo ke zřetelnému odlivu historických témat v Rozpravách ČAVU po tomto roce.

Vizualizace celkových výsledků pomocí knihovny LDAvis pro R (graf č. 6 na s. 80) ukazuje blízkost či vzdálenost jednotlivých témat na dvou osách (jedná o metodu analýzy dat na základě redukce rozměrů). Pokud bych použil heraldickou terminologii, graf je možné šikmo (v protikladu ke kosmo) rozpůlit na dvě poloviny, přičemž horní polovina obsahuje výhradně témata z humanitních a sociálních věd a dolní ze STEM oborů. V kontextu tehdejšího podceňování a podfinancování STEM oborů to svědčí přinejmenším o tom, že abstrahovaná data z publikací jedněch a druhých oborů mají protikladný profil a vědci z každé ze skupin oborů mluví tedy „jiným jazykem“. Vlastně to není nic nového pro dnešní dobu, kdy určité podceňování a podfinancování zaznamenávají naopak humanitní a sociální vědy.

Jediné, co bych autorovi vytkl, je snad až příliš stručné celkové vyhodnocení jeho brilantní analýzy (s. 95-99). Chápu, že v jeho datově orientovaném přístupu připisuje hlavní výpovědní hodnotu samotným datům, resp. výsledkům jejich analýzy, která má (i) exaktní podobu, ale některé rozpoznané vlastnosti dat by asi bylo možné zhodnotit podrobněji. To by si ale vyžádalo mnohem větší rozsah textu kvalifikační práce a i tak mohu jen konstatovat, že student v mnoha ohledech překonal běžné požadavky na diplomovou práci, ať už se jedná o inovativní přístup, ucelenou argumentaci, bohatou diskusi nad některými výsledky a podobně.

Postup využitý v práci (dokumentovaný v příloze č. 2 a kódem v úložišti GitHub, resp. daty v úložišti Zenodo) považuji za inovativní a mimořádně přínosný, a to nejen v českém prostředí. Jednotlivé prvky

analýzy sice využívají dostupné nástroje, ale jejich kombinace a nasazení na data z digitální knihovny za daným cílem je myslím ojedinělé. Proto, pokud je to možné, bych tuto práci rád navrhl na ocenění.

Spolupráci se studentem v průběhu psaní práce hodnotím jako vynikající, diplomant přicházel s vlastními inovativními nápady a pečlivě zohledňoval moje průběžné připomínky k práci.

Podle automatické kontroly v SISu práce nevykazuje znaky plagiátu (tedy neobsahuje pasáže, které by byly shodné s jinými pracemi, resp. detekované shody je možné vysvětlit např. citováním stejných delších titulů knih apod.).

3. Otázky či tematické okruhy k obhajobě

Považujete za pravděpodobné, že přístupy podobné tomu, který jste užil v práci, v budoucnu v knihovnách nahradí dosud ručně prováděné přiřazování předmětových hesel a klíčových slov?

4. Závěrečné zhodnocení

Práci doporučuji k obhajobě a navrhuji ji hodnotit známkou **výborně**.

V Praze dne 18. 1. 2024

doc. Mgr. Jindřich Marek, Ph.D.

Tabulka k hodnocení práce

Bodový zisk za práci	Hodnocení
83 až 100 bodů	Výborně (1)
66 až 82 bodů	Velmi dobře (2)
50 až 81 bodů	Dobře (3)
49 bodů a méně	Neprospěl/a, nedoporučeno k obhajobě