

Posudek oponenta diplomové práce

BARTOŠ, Martin. *Analýza produkce vědy v oborech společenských a humanitních věd v ČR*. Praha: Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví, 2021. Vedoucí diplomové práce PhDr. Adéla Jarolímková, Ph.D.

Předkládaná diplomová práce představuje datové zdroje a možný postup kvantitativní analýzy produkce SHV v institucích v ČR.

K obsahu práce

Úvod je obsáhlým a vcelku zdařilým uvedením do problematiky, ve které se stýká vymezení společenských a humanitních věd, bibliometrie včetně datových zdrojů a účelů a postupů hodnocení vědy. V oddíle 1.2 (Vymezení pojmů) se drží definic zavedených českou legislativou a v případě vymezení typů výsledků RIV zabíhá do přílišné detailnosti. Vhodnější by zde možná bylo odvolat se na Frascati manual z produkce OECD a výčet druhů výsledků národního systému RIV umístit do přílohy.

V dalších kapitolách jsou rozebírány dostupné zdroje údajů o výsledcích VaVal, a to jak na úrovni národní (IS VaVal), tak institucionální. Některé skutečnosti však autor interpretuje nesprávně. Např. tvrdí-li, že „[d]o RIV se tak nedostanou populárně naučné výsledky, články v denním tisku, učební materiály, abstrakta či závěrečné vysokoškolské práce, pakliže nevznikly se státní podporou“ (oddíl 2.2.1), nemá zcela pravdu. Všechny uvedené typy výstupů mohou být do RIV předloženy, avšak jako výsledky typu O (ostatní), které obsahují jen minimálně strukturovaný metadatový záznam a do různých hodnocení vstupují jen velmi sporadicky. V praxi pak některé instituce při zvážení výsledného efektu v poměru k nutnému úsilí výsledky těchto druhů do RIV nepředkládají, leda by je k tomu vedl zvláštní závazek např. ze smlouvy o poskytnutí podpory. Ale dokonce i pokud výsledky vznikly bez veřejné podpory (nikoli nutně státní), mohou být do RIV výzkumnými organizacemi (nikoli však jinými předkladateli) vykázány.

V oddílu 4.1 (Proces posuzování použitelnosti dat) je používán pojem *entita*, z kontextu je však zřejmé, že daleko spíše jde o *atribut* ve smyslu datového modelování a o *atribut* či *proměnnou* ve smyslu statistickém. Při posuzování vyplněnosti toho kterého atributu jsou zobrazovány roky publikace výsledku (RIV: tzv. rok uplatnění výsledku), což však nezohledňuje čas, kdy byl daný atribut do posuzovaného informačního systému přidán, neboť ve všech případech je umožněn dodatečný sběr údajů o výsledcích publikovaných v minulosti starší než je předešlý rok. Proto např. perzistentní identifikátor DOI je inzerován jako přítomný v RIV od roku 1992, ač reálně byl do RIV přidán až v roce sběru 2012). Autor takto sám sebe přesvědčil, že v RIV je druh přístupu k plnému textu výsledku k dispozici již od roku 2000, tři roky před vydáním Berlínské deklarace Open Access. Reálně však tento údaj byl přidán teprve v roce 2020, což vrací úvahy o bleskové adopci Open Access v ČR zpět na zem. Obdobně v oddílu 4.1.1.1 se nachází tvrzení, že „lze potvrdit počáteční rok sběru výsledků v RIV (1991)“, ale zde měl autor opět na mysli rok uplatnění. První sběr dat do RIV přece nemohl proběhnout dříve, než tato komponenta IS VaVal vznikla, tj. v roce 1998 – což autor uvádí v oddílu 2.2.1.

Zároveň autor při analýze “použitelnosti dat” nezohledňuje, že některé údaje mají smysl jen pro některé typy výsledků. Například číslo patentu se týká patentů a užitečných a průmyslových vzorů, které ovšem tvoří méně než 1 % všech výsledků. A tak přestože je tento údaj pro relevantní

výsledky v RIV vyplněn všude, z pohledu Tabulky 6 je prezentován jako nevyplněný pro 99% výsledků. Směrodatnější by však byla informace, že údaj je vyplněný pro 100% výsledků, pro které mohl a měl být vyplněn. Tento do jisté míry extrémní příklad se opakuje pro většinu atributů. Tabulku 6 proto pro posouzení úplnosti dat nelze považovat za užitečnou.

Naopak pozitivně je třeba hodnotit způsob, kterým se autor vyrovnal s přechodem RIV (a návazně i institucionálních systémů) z původní oborové klasifikace na mezinárodní standardní klasifikaci FORD od roku sběru 2018. Jedině vynechání výsledků CERGE-EI (společné pracoviště Univerzity Karlovy a Národohospodářského ústavu AV ČR) z analýzy obsahu institucionálních systémů v oboru ekonomie působí dojem poněkud hrubého zacházení, které může, vzhledem k významu toho pracoviště, výsledky významněji zkreslit. Autor zde mohl zjistit podíl výsledků vykázaných do RIV z UK a z AV ČR a zjistit relativní velikost jejich průniku.

Prezentace výsledků zpracování relativně rozsáhlých dat v přílohách je také zdařilá, kombinuje grafickou a tabulkovou formu. Je vidět, že této prezentaci bylo věnováno značné úsilí. Závěry, které autor vyvozuje, jsou v obecnosti platné, ovšem je nutno přihlížet k výše uvedeným dílčím nedostatkům.

Ke zpracování práce

Práce je psána srozumitelně, občas se vyskytují gramatické chyby nebo nespisovné výrazy. Celkově dobrý dojem narušují některé křížové odkazy na jiné sekce textu, které před vlastním číslem sekce obsahují ještě 0.0 (v Úvodu i jinde). Nejde však o překážku srozumitelnosti.

Ke výsledkům nástrojů pro detekci pokusů o plagiátorství

Systém Theses.cz našel dokumenty se shodou v definici výzkumu a dalších pojmů dle zákona č. 130/2002 Sb., který však autor řádně odkázal. Rovněž našel shody v položkách číselníku oborů, typů výsledků a v názvech vědeckých institucí v ČR, které také nejsou relevantní. Další shody nalezeny v seznamu literatury, také nerelevantní.

Systém Turnitin našel jen útržkovité shody, a to buď tam, kde práce vždy uvádí pramen, nebo u zcela obecných formulací. Většina shod se koncentruje v seznamu literatury, kde však je jejich výskyt přirozený.

Nástroje tak nenalezly žádné relevantní známky plagiátorství.

Otázky k obhajobě

Jak by uchazeč charakterizoval rok sběru dat a rok uplatnění v RIV? Jaký je mezi nimi rozdíl? Jaká nerovnost mezi nimi platí?

Celkové zhodnocení

Práci celkově hodnotím jako v zásadě solidní, v některých ohledech velmi dobrou. Má některé dílčí nedostatky, které se však nedotýkají přímo jádra zkoumaného problému. **Navrhuji známku dobře.**

Hodnocení diplomové práce

Aspekt kvalifikační práce	Vysvětlení	Možné hodnocení	Hodnocení oponentem práce
metodologie a věcné zpracování tématu	Práce představuje problematiku většinou správně a zdařile, ale obsahuje některé faktické nesrovnalosti (které však nezasahují přímo jádro zkoumané problematiky).	0-40 bodů	14 bodů
přínos a novost práce	Práce je přínosná, avšak ponechává relativně velký nevyužitý potenciál.	0-20 bodů	15 bodů
citování, korektnost citování, využití inf. zdrojů	Citování je korektní, práce využívá relevantní zdroje. Formátování odkazů není standardní.	0-20 bodů	14 bodů
slohové zpracování	Práce je psána odborným stylem a srozumitelně.	0-15 bodů	13 bodů
gramatika textu	Gramatické chyby se občas vyskytují.	0-5 bodů	3 body
CELKEM		0-100 bodů	59 bodů