

Bibliometrické zprávy pro výzkumné organizace

Popis statistik pro analýzu výsledků dle Web of Science

Úvod

V rámci hodnocení výzkumných organizací podle *Metodiky 2017+* jsou každé výzkumné organizaci zasílány bibliometrické podklady zpracované *Odborem Rady pro výzkum, vývoj a inovace* na základě materiálu *Bibliometrická analýza pro hodnocení v roce 2017* schváleném na 327. zasedání Rady pro výzkum, vývoj a inovace dne 30. června 2017.

Hlavním smyslem podkladů je poskytnout pohled na oborovou strukturu výsledků a jejich rozložení v kvalitativních pásmech daných bibliometrickými ukazateli vycházejícími z databáze Web of Science.

Bibliometrická analýza byla navržena jednotně pro celý výzkumný „ekosystém“ zahrnující instituce nejrůznějšího zaměření a má tak pochopitelně pro různé organizace různou vypovídací hodnotu. Relevance bibliometrie je daná jak obory, ve kterých organizace publikuje (míra korelace vědecké kvality a výsledků bibliometrie je v různých oborech různá), tak i jejím zaměřením (například v oblasti aplikovaného výzkumu je obvykle relevance bibliometrie nízká). Pro výzkumné instituce, pro které není publikování v impaktovaných časopisech relevantním měřítkem kvality, jsou poskytnuté údaje **pouze doplňkovým kritériem hodnocení**. Proto mají jednotlivé organizace možnost (nikoliv povinnost) se ke zpracovaným podkladům vyjádřit. Hlavním smyslem těchto komentářů je zasadit bibliometrickou analýzu do specifického kontextu konkrétní výzkumné organizace zohledňující například její oborová specifika, orientace na jiné publikační nebo nepublikační výsledky apod.

Širší rámec pro dodané podklady poskytují **bibliometrické zprávy pro jednotlivé obory** resp. oborové skupiny, které byly zpracovány *Odborem Rady pro výzkum, vývoj a inovace* a komentovány členy/členkami oborových panelů ustanovených pro hodnocení podle *Metodiky* v roce 2017. Smyslem oborových komentářů je zasadit statistiky do příslušného oborového národního i mezinárodního kontextu a poskytnout zpětnou vazbu pro další rozvoj nástrojů bibliometrické analýzy v následujících letech. Obdobnou roli pro rozvoj analýz mají též komentáře poskytnuté výzkumnými organizacemi.

Každoročně zpracovávané bibliometrické zprávy mají charakter monitoringu, proto mají relativně jednoduchou podobu. Jednou za pět let, během kompletního hodnocení za uplynulé období, budou použity i další bibliometrické ukazatele (vycházející například z citací).

1 Použité ukazatele

Article Influence Score (AIS) reflektuje průměrný vliv článků v určitém časopisu v průběhu prvních pěti let po jejich zveřejnění v rámci databáze Web of Science (WoS). Odvozuje se z ukazatele Eigenfactor, vycházejícího z toho, kolikrát byly články z daného časopisu zveřejněny v posledních pěti letech citovány v databázi časopisů Journal Citation Reports (JCR), a odpovídá tedy zhruba pětiletému impakt faktoru časopisu (5-Year Journal Impact Factor). Eigenfactor ale navíc zohledňuje kvalitu časopisů, ze kterých tyto citace pocházejí – vysoce citované časopisy mají větší váhu než časopisy méně citované. Odkazy z jednoho článku v časopisu na jiný článek v témže časopisu se nezapočítávají. Hodnota AIS pak udává vliv časopisu normalizovaný v rámci databáze JCR. Hodnoty Eigenfactor (a tedy i AIS) jsou úměrné počtu článků publikovaných v daném časopise – pokud dva časopisy publikují

stejně články se stejným počtem citací a první jich publikuje dvakrát více, pak je i jeho Eigenfactor dvakrát větší než u druhého časopisu.¹

Výše ukazatele AIS slouží k rozdělení časopisů (respektive výsledků) do pásem. Pásma jsou definována pomocí **prvního decilu** a **kvartilů**, dalším ukazatelem je **medián**. Pásma jsou vytvářena vždy **pro každý obor zvlášť**.

Kvartily rozdělují **pořadí časopisů** na čtvrtiny (tj. první kvartil zahrnuje 25 % nejlepších časopisů dle AIS, poslední čtvrtý kvartil naopak zahrnuje spodních 25 % časopisů). **První decil** AIS tvoří 10 % nejlepších časopisů v oboru.²

Dalším použitým ukazatelem je **medián**. Medián udává „střední“ hodnotu dané veličiny, ale na rozdíl od průměru není citlivý na extrémní minimální a maximální hodnoty. Je určen jako hodnota AIS, která se nalézá uprostřed pořadí.³ Jinými slovy medián lze též chápat jako hranici mezi druhým a třetím kvantilem.

Analýza využívá pro tvorbu kvartilů, decilu a mediánu nejen pořadí časopisů, ale též **oborové pořadí článků**. Pořadí je vytvořeno tak, že se všechny celosvětově publikované články za daný rok v daném oboru seřadí podle AIS časopisu, ve kterém jsou publikovány. Do pořadí jsou zařazeny pouze citovatelné dokumenty (citable items). Tento přístup zohledňuje celkové počty článků v jednotlivých kvalitativních pásmech a dovoluje tudíž vzájemně porovnat podíly národních a s mezinárodních výsledků v jednotlivých pásmech.⁴

2 Charakteristika dat

2.1 Oborová struktura

Data z Web of Science, oborově strukturovaná do *WoS categories fields* (WC), jsou převedena do **oborové struktury dle OECD: Fields of Research and Development** (FORD). Převod se provádí na základě materiálu *Struktura oborů OECD (Frascati Manual) – převodník M17* schváleném RVVI⁵.

Oborová struktura OECD je užší (41 FORD), oborová struktura Web of Science je zhruba šestkrát větší (252 WC). Klastrování oborů do širších FORD kategorií poskytuje tedy oborový pohled na obecnější úrovni. To dovoluje hromadně zpracovat výstupy hodnocení v přehlednější struktuře a zároveň ustavit racionálnější strukturu panelů i počet panelistů.

V případě **multioborových časopisů** se jejich AIS používá při analýzách všech oborů, ve kterých jsou zařazeny.⁶

2.2 Propojení databází IS VaVaI a Web of Science

Propojení údajů z RIV s daty Web of Science probíhalo primárně na základě unikátního kódu každé publikace (*accession number*). Protože toto pole bylo do konce roku 2017 v RIV nepovinné, byly

¹ Viz www.eigenfactor.org/about.php.

² Protože je první decil součástí horního kvartilu, jsou výsledky prvního decilu součástí prvního kvartilu a součty podílů za všechny kvartily a první decil tedy dávají více než 100%.

³ Pokud má soubor sudý počet prvků, mediánem je průměr dvou prostředních sousedících hodnot.

⁴ Rozložení počtu článků v kvantilech podle AIS časopisu není rovnoměrné, v některých oborech se např. v I. kvantilu publikuje 35 % všech článků daného oboru.

⁵ 324. zasedání Rady pro výzkum, vývoj a inovace dne 31. března 2017.

⁶ Výjimkou je níže popsaný výstup VO2a a VO2b kumulující všechny výsledky za danou výzkumnou organizaci bez ohledu na oborovou příslušnost výsledku.

výsledky dále napojovány pomocí ISSN/e-ISSN/ISBN a názvů.⁷ Tento postup je ovlivněný nekorektními záznamy v RIV. Při tvorbě podkladů byla proto řada chybných záznamů v matici opravena za účelem zahrnutí co největšího počtu výsledků do analýzy. Avšak objem výsledků nedovoluje provést úplnou korekci dat. Chyba má náhodný charakter a pravděpodobně se po provedené opravě týká zlomku analyzovaných dat.

Protože databáze Web of Science i Scopus jsou živými, průběžně aktualizovanými databázemi, část výsledků v RIV nemusela být propojena nebo doplněna o ukazatel AIS, přestože lze tyto údaje v současné chvíli individuálně dohledat. Platí proto, že data, která bude **možno propojit zpětně**, budou do analýzy zahrnuta vždy v následujícím roce.

2.3 Spolupráce výzkumných organizací

Metodika 17+ u výsledků vzniklých ve spolupráci již nezohledňuje jednotlivé podíly, daný výsledek je započten každé organizaci, která se podílela na jeho tvorbě.⁸ V příštím roce budou specificky identifikovány výsledky vzniklé v rámci velkých autorských týmů, u nichž bývá příspěvek spoluautorů víceméně okrajový. Problematika vyhodnocování výsledků vzniklých ve spolupráci bude diskutována a dále rozpracována.

2.4 O čem předložené statistiky informují a o čem neinformují

Předložená data striktně vzata vypovídají pouze o distribuci článků výzkumných organizací v impaktovaných časopisech rozdělených podle jejich významu z hlediska citovanosti do zvolených pásem. Míra, do jaké tento přístup vypovídá o kvalitě jednotlivých výstupů a produktivitě jednotlivých výzkumných organizací, se liší v závislosti na jejich zaměření a na publikačních zvyklostech konkrétních oborů, ve kterých výzkumná organizace publikuje.

3 Popis výstupů

3.1 Bibliometrická zpráva za výzkumnou organizaci

Podklad VO1: seznamy výsledků výzkumné organizace a jejich oborové členění.

Z důvodů rozsahu jsou tabulky VO1a1 - VO1a4 dodávány v odděleném souboru ve formátu MS Excel obsahujícím následující čtyři listy:

VO1a.1 - výsledky VO obsahuje seznam výsledků dané organizace seřazených sestupně podle výše AIS.

VO1a.2 – neanalyzované je doplňkový list obsahující souhrn neanalyzovaných výsledků, které byly propojeny s databází WoS, avšak nemají přidělen AIS, respektive daný časopis tímto ukazatelem nedisponoval.⁹ Další množinu tvoří výsledky, které nebyly propojeny například z důvodu změny ISSN časopisu nebo chybného záznamu v RIV. Všechny tyto výsledky budou znovu propojovány v následujícím roce.

VO1a.3 - I. kvartil-časopisy obsahuje výsledky organizace zařazené do I. kvartilu podle pořadí časopisů s vyznačením výsledků, které jsou navíc v I. decilu. Uvádí i celkový počet takto zařazených výsledků. V případě, že je časopis zařazen do více oborů, započítává se obor s nejvyšším umístěním.

⁷ Vyplňovací povinnost pro databáze Scopus a Web of Science platí od 1. 1. 2018.

⁸ Pro úroveň oborových analýz jsou naopak data deduplikována a každý výsledek vstupuje do oborové analýzy logicky pouze jednou.

⁹ Obvyklým důvodem je, že takový časopis zanikl, nebo existuje příliš krátkou dobu. Databáze Web of Science údaje doplňuje průběžně, proto je možné, že danému časopisu byl či bude ukazatel doplněn, avšak v datových podkladech dodaných Clarivate Analytics, vlastníka Web of Science, nebyl evidován.

VO1a.4 - I. kvartil-články obsahuje stejnou informaci jako předchozí tabulka, ale založenou na pořadí článků zohledňující celkovou celosvětovou produkci a distribuci textů v jednotlivých pásmech.

Tabulka a graf VO1b: udává počet a podíl výsledků organizace podle oborové příslušnosti časopisů, ve kterých byly zveřejněny.

Podklad VO2: základní bibliometrická charakteristika VO

Tabulka a graf VO2a: udává souhrnné rozložení všech výsledků výzkumné organizace v prvním decilu a v kvartilech podle mezinárodního pořadí časopisů. V případě, že je časopis výsledku zařazen do více oborů, je výsledku přiřazena nejvyšší dosažená hodnota.

Tabulka a graf VO2b: stejný údaj jako v předchozí tabulce, ale podle mezinárodního pořadí článků.

Podklad VO3: excelentní výsledky podle AIS.

Tabulka VO3: udává počet nejlepších výsledků výzkumné organizace podle umístění v prvním decilu a prvním kvartilu podle pořadí časopisů i článků rozčleněný podle oborů. Zároveň udává procentní podíl těchto výsledků na všech výsledcích výzkumné organizace v daném oboru. V tabulce jsou započteny všechny obory, ve kterých je výsledek (resp. příslušný časopis) indexován. Obecně se tedy jeden výsledek může vyskytnout ve více oborech.

Podklad VO4: oborové srovnání výsledků VO se světem a s ČR.

Tabulka VO4a: udává bibliometrický profil výzkumné organizace podle oborů: rozložení výsledků v prvním decilu a v kvartilech podle mezinárodního pořadí časopisů. Kromě počtu výsledků v daném oboru a percentilu udává jejich procentní podíl na všech výsledcích organizace v daném oboru.

Grafy VO4b: tabulka porovnává procentuální rozložení výsledků výzkumné organizace v kvartilech a prvním decilu s mezinárodním rozložením v daném oboru. *Do porovnání jsou zahrnuty pouze obory, ve kterých má výzkumná organizace alespoň 10 výsledků.*

Tabulka a graf VO4c: Porovnání mezinárodního oborového mediánu pořadí článků s oborovými mediány výzkumné organizace. *Do porovnání jsou zahrnuty pouze obory, ve kterých má organizace alespoň 10 výsledků.* Tabulka agreguje informaci z předchozího grafu.

Tabulka a graf VO4d: Podobné porovnání jako v předchozí tabulce, ale medián výzkumné organizace se porovnává s národním oborovým mediánem.