

Sergiusz Łuczak

Politechnika Warszawska
<https://orcid.org/0000-0003-2351-9473>
sergiusz.luczak@pw.edu.pl

Wioletta Jachym

Państwowa Wyższa Szkoła Zawodowa w Tarnowie
<https://orcid.org/0000-0002-1153-5727>

Jakość informacji w bibliograficznych bazach danych

Information Quality in Bibliographic Databases

Abstrakt

Celem artykułu jest analiza jakości informacji zawartych w bibliograficznych bazach danych Web of Science oraz Scopus. W wyniku przeprowadzonych badań wskazano, iż z powodu błędów występujących w analizowanych bazach często zdarzają się przypadki znacznego zaniżenia wartości parametrów bibliometrycznych. W ekstremalnych przypadkach błąd względny danego parametru przekracza kilkadziesiąt procent – dotyczy to zarówno indywidualnych autorów, jak i całych czasopism naukowych. Niniejsze opracowanie przedstawia omówienie źródeł błędów w bibliograficznych bazach danych oraz metody ich efektywnego wyszukiwania i korekty.

Słowa kluczowe

cytowanie, Impact Factor, CiteScore, Web of Science, Scopus

Abstract

The aim of the article is to analyze the quality of information contained in the bibliographic databases Web of Science and Scopus. The results of the presented study prove that due to

Informacja o artykule / Article Information

Otrzymano (Received): 23.11.2021 • Przyjęto do druku (Accepted): 1.12.2021 • Opublikowano (Published): Grudzień (December) 2021

errors existing in the analyzed databases, there are often cases of significant underestimation of bibliometric parameters. In extreme cases, the relative error of a given parameter exceeds several dozen per cent, which concerns both individual authors and entire scientific journals. The publication presents a discussion on sources of errors in bibliographic databases and methods of their effective search and correction.

Keywords

Citation, Impact Factor, CiteScore, Web of Science, Scopus

Wstęp

Podstawowym celem nauki jest tworzenie i prezentowanie wiedzy, a główną funkcją społeczną jest funkcja poznawcza. Niestety obraz ten został zdeformowany poprzez zjawisko „biurokratyzacji nauki polskiej i życia akademickiego”. Alicja Żywczok podkreśla, że „na uniwersytetach zapanowała gorączkowa dyskusja już nie na temat fascynującej problematyki badań naukowych czy dochodzenia do poznania prawdy, ale drobiazgowej sprawozdawczości, która powoduje, że uwaga i energia pracowników skierowana jest na gromadzenie jak największej liczby publikacji mających najwyższą punktację”¹. W potocznym użyciu funkcjonują terminy „punktozy”, „słotozy”, „gry parametrycznej”, „grantozy”, które zostały już wielokrotnie zinterpretowane². Problem jest wciąż aktualny w środowisku naukowym, zwłaszcza w kontekście zbliżającej się ewaluacji jakości działalności naukowej według nowych zasad w 2022 roku.

W podsumowaniu debaty dotyczącej ewaluacji nauki, która odbyła się 25 lutego 2021 roku i przeprowadzonej przez organizację Inicjatywa Wolna Nauka, przedstawiono problemy i zagrożenia obecnego systemu ewaluacji, wymieniono m.in.³:

- pisanie artykułów „pod wydawnictwo i czasopismo”;
- zjawisko publikowania „pod stopień naukowy”;

¹ A. Żywczok, *Współczesny uniwersytet – w pułapce racjonalności eksplikatywnej*, „Rocznik Lubuski” 2018, t. 44, cz. 2, s. 94.

² Por. m.in.: T. Szczerbowski, *Punktoza jako słowo ostatnich lat*, „Poradnik Językowy” 2017, z. 7, s. 80–87; W. Włoskiewicz, *Nowomowa pseudometanaukowa? O języku polskiej polityki naukowej*, „Poradnik Językowy” 2019, z. 10, s. 35–47; E. Kulczycki, *Punktoza jako strategia w grze parametrycznej w Polsce*, „Nauka i Szkolnictwo Wyższe” 2017, nr 1(49), s. 63–78; K. Kulikowski, E. Antipow, *Niezamierzone konsekwencje punktozy jako wartości kulturowej polskiej społeczności akademickiej*, „Studia Socjologiczne” 2020, nr 3(238), s. 207–236.

³ Inicjatywa Wolna Nauka, *Debata o ewaluacji nauki: podsumowanie*, <https://www.wolnanauka.org/debata-o-ewaluacji-nauki-podsumowanie/> [dostęp: 12 sierpnia 2021 r.].

- dopisywanie przełożonych do artykułów;
- wstrzymywanie recenzji przez recenzentów;
- spadek jakości badań naukowych;
- ograniczenie możliwości replikacji badań i publikacji ich wyników;
- pęd do nowości i oryginalności (fałszywie rozumianych), czego konsekwencją jest pogarszanie się jakości pracy naukowców.

Drugim, niemniej ważnym elementem hamującym rozwój polskiej nauki jest dominująca pozycja dwóch bibliograficznych baz danych – Web of Science oraz Scopus, które pośrednio związane są z rozwojem naukowego dorobku indywidualnego i instytucjonalnego nauki polskiej oraz z ważnymi procesami decyzyjnymi, w tym: przyznawaniem stopni i tytułów naukowych⁴, czy wymaganiami stawianymi przed wnioskodawcą przy udzielaniu finansowania na realizację i raportowanie grantów naukowych⁵. Dodatkowo można podkreślić również dyskryminację w tych bazach nauk humanistyczno-społecznych⁶, tekstów pisanych w językach narodowych oraz brak wydawnictw o charakterze monograficznym.

Niestety bazy te nie zawierają informacji o najwyższej jakości⁷. W wyniku przeprowadzonych badań wskazano, iż z powodu błędów występujących w przywołanych bazach często zdarzają się przypadki znacznego zaniżenia wartości parametrów bibliometrycznych. W związku z powyższym w przedstawionym opracowaniu zwrócono szczególną uwagę na jakość informacji zawartych w WoS i Scopus oraz możliwość ich wykorzystania.

Z uwagi na fakt, iż informacja jest obecnie coraz bardziej wartościowym zasobem, to właśnie jej jakość ma kluczowe znaczenie w podejmowaniu ważnych decyzji oraz w całym procesie zarządzania informacją. Terminu jakości informacji nie można przedstawić w postaci uniwersalnej definicji, gdyż dla różnych odbiorców może oznaczać zupełnie

⁴ Zob. Rada Doskonałości Naukowej, *Wymagania dokumentacyjne wniosków w sprawie nadania tytułu profesora*, <https://www.rdn.gov.pl/dl/219/attachment/68d7c4/4%20Wykaz%20opublikowanych%20prac%20naukowych%20lub%20tw%C3%B3rczych%20prac%20zawodowyc.docx> [dostęp: 12 sierpnia 2021 r.].

⁵ Zob. Narodowe Centrum Nauki, *Raportowanie projektów NCN. Raport końcowy – wykaz publikacji*, <https://www.ncn.gov.pl/finansowanie-nauki/informacje-dla-realizujacych-projekty/raporty-ncn/koncowy/publikacje> [dostęp: 12 sierpnia 2021 r.].

⁶ Por. m.in.: V.K. Singh, et al., *The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis*, „Scientometrics” 2021, no. 126, s. 5113–5142; M.A. Vera-Baceta, M. Thelwall, K. Kousha, *Web of Science and Scopus language coverage*, „Scientometrics” 2019, no. 121, s. 1803–1813; P. Mongeon, A. Paul-Hus, *The journal coverage of Web of Science and Scopus: a comparative analysis*, „Scientometrics” 2016, no.106, s. 213–228.

⁷ Zob. F. Franceschini, D. Maisano, L. Mastrogiacomo, *Empirical analysis and classification of database errors in Scopus and Web of Science*, „Journal of Informetrics” 2016, no. 10.4, s. 933–953.

coś innego. Na podstawie literatury przedmiotu można natomiast zidentyfikować pewne atrybuty informacji odnoszące się do jej jakości. Wiesław Babik koncentrując się na jakości informacji, zwraca uwagę na „ogół własności obiektu wiążących się z jego zdolnością do zaspokojenia oczekiwanych potrzeb”⁸. Jakość informacji to stopień usatysfakcjonowania odbiorcy m.in. użytecznością, poprawnością i aktualnością⁹. Informacja ma więc znaczenie dla użytkownika, jeżeli umożliwia mu zaspokojenie jego potrzeb informacyjnych. Definicją obejmującą szerokie spektrum oczekiwanych atrybutów informacji jest sformułowanie wskazujące, iż jakość informacji to „suma własności i wymiarów informacji pozwalająca zapewnić lub przekroczyć oczekiwania i wymagania oraz wyrażone i nieujawnione potrzeby użytkownika”¹⁰.

Podejmując próbę wyszczególnienia cech jakości informacji, można przytoczyć atrybuty najczęściej oczekiwane i wymieniane w literaturze¹¹ – są to m.in.: kompletność, relewantność, użyteczność, poprawność, aktualność, wiarygodność, dostępność. Biorąc pod uwagę wymienione atrybuty, w niniejszym opracowaniu skoncentrowano uwagę na jakości dwóch najbardziej rozpoznawalnych bibliograficznych baz danych – Web of Science oraz Scopus.

Web of Science to jedna z najstarszych naukowych baz danych, stworzona początkowo jako narzędzie do wyszukiwania informacji w 1964 roku przez Eugene Garfielda z Institute of Scientific Information (ISI) z Filadelfii (USA). Baza nosiła nazwę Science Citation Index (SCI), początkowo obejmowała 700 czasopism i miała być przede wszystkim indeksem cytowań. W kolejnych latach została wzbogacona o kolejne indeksy cytowań: Social Sciences Citation Index (SSCI) w 1973 r., Arts & Humanities Citation Index (AHCI) w 1978 r. i Book Citation Index (BKCI) w roku 2011. Indeksy SCI, SSCI i AHCI zostały połączone i uruchomione w sieci WWW, jako Web of Science w 1997 r. Kolejny indeks cytowań o nazwie Emerging Sources Citation Index (ESCI) został uruchomiony w 2015 roku. Obecnie Web of Science jest własnością Clarivate Analytics (dawniej Thomson Reuters)¹². Kolekcja Web of Science (według danych z 12.08.2021 r.) indeksuje ponad 21.100 recenzowanych,

⁸ W. Babik, *Ekologia informacji*, Kraków 2014, s. 31.

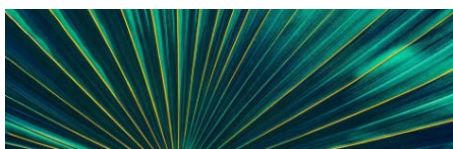
⁹ Zob. B. Stefanowicz, *Informacja*, Warszawa 2004, s. 99–100.

¹⁰ J. Ruževičius, A. Gedminaitė, *Business information quality and its assessment*, „Engineering Economics” 2007, vol. 52.2, s. 19.

¹¹ Zob. Ł. Grudzień, *Koncepcja oceny jakości informacji o procesach w systemach zarządzania*, [w:] *Innowacje w zarządzaniu i inżynierii produkcji*, red. R. Knosala, Opole 2012, s. 638.

¹² Zob. V.K. Singh, et al., *The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis*, op. cit., s. 5115.

wysokiej jakości czasopism naukowych z całego świata, reprezentujących 254 dyscypliny naukowe (w tym czasopisma wydawane w modelu Open Access)¹³. Dodatkowo do bazy jest narzędzie InCites¹⁴, przy użyciu którego można przeprowadzić ewaluację aktywności naukowej danej jednostki, opracować strategie badawcze oraz podejmować decyzje publikacyjne i kadrowo-administracyjne (rys. 1).



- ✔ 1.5 billion cited references dating back to 1900
- ✔ 74.8 million total records
- ✔ 10.1 million total Open Access records
- ✔ 21,100+ unique global journals
- ✔ 254 disciplines

Rys. 1. Baza Web of Science w liczbach

Źródło: Clarivate, Web of Science Core Collection, <https://clarivate.com/webofscience-group/solutions/web-of-science-core-collection/> [dostęp: 12.08.2021 r.].

Baza Scopus jest bazą bibliograficzną, indeksującą wyłącznie recenzowane publikacje naukowe, doniesienia konferencyjne, książki, a także patenty. Na liście ponad 7000 wydawców znajdują się w niej m.in.: Elsevier, Springer, Wiley-Blackwell, Taylor & Francis, Sage, Wolters Kluwer, Oxford University Press, Emerald, Inderscience Publishers, Cambridge University Press, Bentham Science, IEEE. Scopus jest interdyscyplinarną bazą abstraktów i cytowań z zakresu nauk matematyczno-przyrodniczych, technicznych, medycznych, społecznych i humanistycznych. Baza obejmuje ponad 25 000 recenzowanych czasopism (w tym także w modelu Open Access), 234 000 książek, około 8 mln referatów konferencyjnych (rys. 2)¹⁵. W roku 2021 baza zawierała ponad 82 mln rekordów bibliograficznych. W bazie Scopus dostępne jest, analogiczne w stosunku do InCites, narzędzie SciVal¹⁶, przeznaczone do

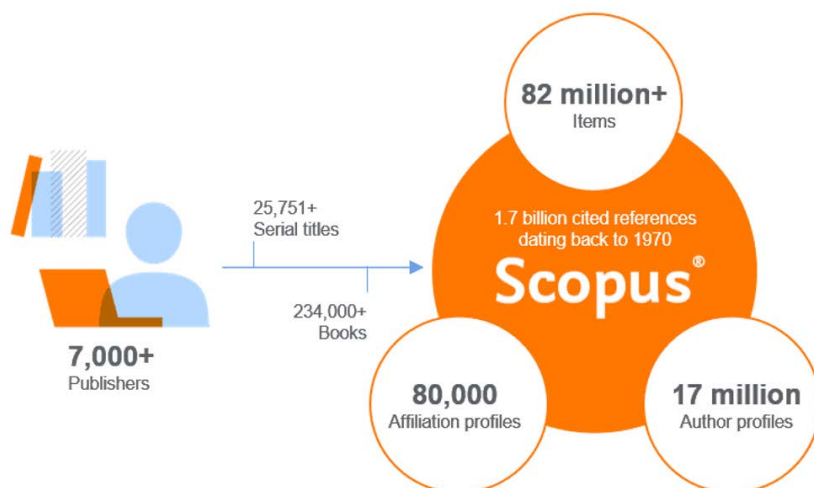
¹³ Zob. Clarivate, *Web of Science Core Collection*, <https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/> [dostęp: 12 sierpnia 2021 r.].

¹⁴ Zob. Clarivate, *InCites*, <https://incites.clarivate.com/> [dostęp: 12 sierpnia 2021 r.].

¹⁵ Zob. Elsevier, *Scopus*, https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030 [dostęp: 12 sierpnia 2021 r.].

¹⁶ Zob. Elsevier, *Czym jest SciVal*, <https://www.elsevier.com/pl-pl/solutions/scival/scival-polska> [dostęp: 12 sierpnia 2021 r.].

analizy danych zawartych w bazie Scopus, w tym wizualizacji osiągnięć instytucji naukowych, porównania z innymi jednostkami, oceny potencjalnych współpracowników oraz partnerów, a także badania trendów w świecie nauki.



Rys. 2. Baza Scopus w liczbach

Źródło: Elsevier, Scopus, https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030 [dostęp: 12.08.2021 r.].

Parametry bibliometryczne w Web of Science oraz Scopus

Odkąd w świecie naukowym za miarę jakości prowadzonych badań naukowych przyjęto liczbę cytowań danej publikacji, toczona jest nieustannie dyskusja i podejmowana krytyka takiego podejścia¹⁷. Wyszukanym jest szereg argumentów za i przeciw takiej metodyce. Niemniej jednak należy stwierdzić, że jak dotąd nie zaproponowano lepszej metody, która byłaby obiektywna i bazowała na analizie danych liczbowych. Rozwój technik komputerowych umożliwia coraz bardziej zaawansowaną analizę i przetwarzanie danych związanych z publikacjami naukowymi, stale wprowadzane są nowe parametry, uwzględniające już nie tylko liczbę cytowań danej publikacji, ale także kontekst czasowy (tzn. „wiek” publikacji), czy dziedzinę wiedzy. Konkurencja pomiędzy bazą Web of Science a Scopus stale zmusza właścicieli tych baz do po-

¹⁷ R. Z. Morawski, *Nauka w czasach biurokracji*, [w:] *Inżynier z duszą humanisty*, red. K. Baranowski, J. Sośnicka, Łódź 2017, s. 321–348.

prawy jakości swoich usług: udoskonalane są interfejsy użytkownika, zwiększa się transparentność metod wyznaczania poszczególnych parametrów bibliometrycznych, ulepszone są interfejsy programowe (API), przy jednoczesnym zwiększaniu zakresu ich funkcjonalności. Natomiast wiele do życzenia pozostawia nadal duża liczba błędów pojawiających się w tych bazach.

Parametry bibliometryczne dla indywidualnych naukowców

Podstawowe parametry bibliometryczne dla indywidualnego naukowca to liczba jego publikacji indeksowanych w danej bazie oraz liczba cytowań każdej z tych publikacji. W przypadku istnienia kilku publikacji w bazie, oblicza się dodatkowo indeks Hirscha (*h*-indeks), który odpowiada maksymalnej liczbie publikacji autora *N*, z których każda została zacytowana co najmniej *N* razy. Dodatkowo uwzględnia się fakt występowania autocytowań: zarówno w przypadku podawania całkowitej liczby cytowań jak i indeksu Hirscha. W 2019 r. baza Scopus zaproponowała jeszcze kilka nowych parametrów bibliometrycznych dla pojedynczego autora – są one dostępne na platformie SciVal i są nazywane *Sowball Metrics*. Te parametry to: 5-letni indeks Hirscha – czyli indeks Hirscha wyznaczony tylko dla publikacji autora z ostatnich 5 lat, oraz parametr o nazwie *Field-Weighted-Citation-Index* (FWCI) relatywizujący liczbę cytowań danej publikacji, biorąc pod uwagę to, jak dawno temu została ona opublikowana, jakiego rodzaju jest to publikacja oraz średnią liczbę cytowań w danej dyscyplinie nauki; wartość „1” oznacza średnią liczbę cytowań, wartość powyżej „1” oznacza większą niż średnia liczbę cytowań w danej dyscyplinie i na odwrót. Wartość parametru FWCI zależna jest od wybranego zakresu czasowego analizy bibliometrycznej (np. 3, 5, 10 lat). Podobne parametry wprowadziła baza Web of Science (za pomocą platformy InCites), np. *Category Normalized Citation Impact*, będący odpowiednikiem FWCI.

Parametry bibliometryczne dla czasopism naukowych

W przypadku czasopism mówimy o zbiorze publikacji wydanych w danym roczniku oraz liczbie ich cytowań. Aby ujednoczyć ocenę czasopism znacznie różniących się liczbą publikowanych artykułów, obie bazy wprowadziły parametr będący ułamkiem, którego licznik stanowi liczba cytowań, natomiast mianownik – liczba artykułów wydanych w danym roczniku. Wartość tego parametru podawana jest raz na rok (z opóźnieniem około półrocznym): w połowie czerwca w przypadku bazy Web

of Science i pod końcem maja w przypadku bazy Scopus. Tutaj istnieją jednak pewne różnice pomiędzy bazami Web of Science oraz Scopus.

Baza Web of Science podaje najbardziej znaczący w świecie nauki parametr bibliometryczny dla czasopisma o nazwie *Impact Factor* (IF). Jego wartość na dany rok kalendarzowy wyliczana jest jako liczba wszelkich publikacji opublikowanych w danym roku kalendarzowym (np. 2021), które są rekordami w tej bazie i które cytują publikacje opublikowane w danym czasopiśmie w dwóch poprzednich latach kalendarzowych (np. 2019, 2020), podzielona przez całkowitą liczbę publikacji opublikowanych w danym czasopiśmie w dwóch poprzednich latach kalendarzowych. Całkowita liczba publikacji obejmuje tylko trzy rodzaje publikacji: artykuły badawcze (*article*), artykuły przeglądowe (*review*) oraz referaty konferencyjne (*paper*) – tzn. w tej liczbie nie uwzględnia się: komentarzy, odpowiedzi na komentarze, listów redaktora itp. Natomiast liczba cytowań dotyczy wszystkich rodzajów publikacji. Baza podaje także 5-letni odpowiednik tego parametru: *5-year Impact Factor*, obejmujący cytowania publikacji nie z dwóch poprzednich lat kalendarzowych, lecz pięciu (np. 2016-2020). Wartość parametrów IF i 5-year IF podawana jest z dokładnością do 3. miejsca po przecinku, np. IF 2020: 2,106. W roku 2021 firma Clarivate Analytics, zarządzająca bazą Web of Science, wprowadzała nowy parametr dla czasopism ujętych w zbiorze Emerging Sources Citation Index (ESCI), które dopiero przechodzą proces 3-letniej ewaluacji i nie mają jeszcze parametru IF oraz dla czasopism humanistycznych. Ten nowy parametr został nazwany *Journal Citation Indicator* i będzie wyznaczany dla perspektywy 3- i 5-letniej.

Baza Scopus wyznacza bardzo podobny do IF parametr dla czasopism, o nazwie *CiteScore* (CS). Różnica polega na tym, że obejmuje on inne roczniki. Do roku 2019 włącznie był to analogiczny parametr do IF, z tą tylko różnicą, że dotyczył on publikacji z trzech poprzednich lat kalendarzowych. Natomiast w roku 2020 wprowadzono dwie zmiany: rozszerzono zakres roczników do 4 lat kalendarzowych, włączając w to rocznik bieżący (np. 2018–2021) oraz wzięto pod uwagę publikacje cytujące nie tylko z bieżącego rocznika (np. 2021), ale ze wszystkich 4 roczników (np. 2018–2021). Wartość parametru CS podawana była z dokładnością do 2. miejsca po przecinku, a obecnie tylko 1. miejsca, np. CS 2020: 5,5. Baza podaje dodatkowo dwa pochodne względem CS parametry: SJR oraz SNIP.

Oprócz tego obie bazy podają, w którym kwantylu mieści się dane czasopismo w zbiorze czasopism przypisanych do danej dyscypliny naukowej (czasami kilku dyscyplin) pod względem liczby cytowań opublikowanych w nim artykułów.

Parametry bibliometryczne dla ośrodków naukowych i akademickich

W przypadku ośrodków naukowych i akademickich zbiorcze parametry bibliometryczne odgrywają kluczową rolę przy ustalaniu ich pozycji na różnych światowych i krajowych listach rankingowych. Najważniejsze parametry to liczba publikacji afiliowanych przez dany ośrodek w prestiżowych czasopismach (różnie definiowanych przez różne rankingi), czasami odniesiona do liczby pracowników, a także liczba cytowań publikacji afiliowanych przez dany ośrodek, której wpływ może sięgać nawet 30% całkowitej oceny¹⁸.

Źródła błędów w bazach

W bibliograficznych bazach danych występuje niezliczona liczba rodzajów błędów, zwłaszcza w przypadku starych rekordów. Nie sposób w tym opracowaniu omówić wszystkie ich rodzaje. Niemniej jednak sytuacja ulega stałej poprawie, ponieważ rozwijane są przez cały czas algorytmy sztucznej inteligencji, które umożliwiają automatyczne poprawianie kluczowych błędów występujących w danym rekordzie.

W niniejszej publikacji uwaga skupiona została na najczęściej występujących błędach, które powodują zaniżenie parametrów bibliometrycznych dla poszczególnych interesariuszy. Są to:

- niewprowadzona do bazy publikacja cytowana;
- niewprowadzona do bazy publikacja cytująca;
- brak publikacji cytowanej w spisie pozycji literaturowych publikacji cytującej;
- nierozpoznanie publikacji cytowanej jako rekordu w bazie (brak hiperlinku przy tytule).

W przypadku autorów publikacji i ośrodków akademickich/badawczych jest to jeszcze brak przypisania publikacji cytowanej do profilu danego autora/uczelni. W przypadku czasopism są to dodatkowo:

- błędny rocznik publikacji cytowanej;
- błędny rocznik publikacji cytującej.

¹⁸ Clarivate, *Institute for Scientific Information: An introduction to global university rankings*, <https://clarivate.com/lp/an-introduction-to-global-university-rankings/> [dostęp: 20 listopada 2021 r.].

Ciekawostką jest fakt, że czasami zdarzają się błędy na korzyść danego interesariusza, gdyż mogą się pojawić pojedyncze cytowania w rzeczywistości niewystępujące. Zdarza się to jednak bardzo rzadko.

Wyszukiwanie błędów w bibliograficznych bazach danych

Szczegółowe wskazówki dotyczące efektywnych sposobów wyszukiwania błędów skutkujących zaniżeniem parametrów bibliometrycznych będą przedmiotem kolejnej publikacji autorów. W ramach niniejszej pracy krótko omówione zostały podstawowe zasady wyszukiwania błędów w bazach.

Podstawowa metoda to porównanie ze sobą treści dwóch baz dla danej publikacji (lub zestawu publikacji). Najbardziej efektywna jest analiza cytowań dla danej publikacji w bazie Web of Science oraz Scopus. Należy wyszukać w obu bazach daną publikację (jeśli jest ona oczywiście indeksowana w obu bazach) i porównać ze sobą listę publikacji cytujących ją. Często pojawiają się pewne różnice. Zazwyczaj wynikają one z tego, że obie bazy różnią się nieco zbiorem indeksowanych czasopism i pozycji (np. materiałów konferencyjnych), w związku z czym dana publikacja cytująca może być indeksowana tylko w jednej z baz. Natomiast zdarzają się też przypadki, że publikacja cytująca jest indeksowana w obu bazach, a generowane przez nią cytowanie widoczne jest tylko w jednej z nich. Takich właśnie publikacji cytujących szukamy.

Dalsze rezultaty można osiągnąć, na tej samej zasadzie porównując jedną z tych baz (korzystniej bazę Scopus) z wyszukiwarką Google Scholar. Z tym, że jest to znacznie mniej efektywne, ponieważ Google Scholar nie jest naukową bazą danych, a wyszukiwarką i występuje w niej dużo więcej błędów, zgromadzonych jest dużo więcej publikacji typu prace dyplomowe, prace doktorskie itp., które nie są rekordami w bazie WoS ani Scopus.

Jeszcze inne źródła, z którymi można porównać bazę Scopus lub Web of Science, to serwis ResearchGate (wymagane jest posiadanie swojego konta) albo strona czasopisma/wydawcy, w którym ukazała się dana publikacja, jeśli są tam zamieszczone dane o cytowaniach (np. IEEE Digital Library).

Błędy, które pojawiają się w bibliograficznych bazach danych, powstają czasem już na etapie tworzenia publikacji. Dotyczy to zwłaszcza referatów konferencyjnych, gdy zazwyczaj jest mało czasu na przygotowanie materiału do druku i nikt nie sprawdza poprawności pozycji bibliograficznych. Inną przyczyną są skróty nazw czasopism używane

przez różne wydawnictwa, które mogą powodować niejednoznaczność informacji w przypadku istnienia kilku czasopism o podobnych nazwach. Błędy mogą też powstawać podczas skanowania starszych dokumentów albo przy przetwarzaniu znaków diakrytycznych.

W ciągu kilku lat badań jakości informacji w bibliograficznych bazach danych zebrano dużą liczbę informacji. Zostały one przedstawione w poniższych tabelkach. Dwie pierwsze dotyczą polskich naukowców z różnych ośrodków akademickich (m.in. Politechnika Warszawska, Akademia Górniczo-Hutnicza). W celu zachowania anonimowości, usunięto ostatnie 4 cyfry identyfikatora ORCID. W tabelach przedstawiono ostateczną wartość parametrów bibliometrycznych w bazie Web of Science oraz Scopus, po zgłoszeniu zauważonych błędów i ich korekcie.

Tab. 1. Cytowania oraz indeks Hirscha dla pojedynczych naukowców w bazie Web of Science

Nr ORCID	Ostateczna liczba cytowań	h-indeks	Cytowania odzyskane	Zwiększenie h-indeksu
0000-0003-2351-...	61	4	19,7%	0
0000-0002-2529-...	506	11	2%	1
0000-0002-8650-...	24	2	12,5%	1
0000-0003-3817-...	34	3	5,9%	0
0000-0001-7741-...	134	4	10,4%	0
0000-0002-1049-...	18	3	5,6%	0
0000-0002-6946-...	14	2	21,4%	1
0000-0002-3720-...	58	4	70,7%	2
Średnio:			18,5%	0,6

Źródło: Opracowanie własne na podstawie wyników badań.

Tab. 2. Cytowania oraz indeks Hirscha dla pojedynczych naukowców w bazie Scopus

Nr ORCID	Ostateczna liczba cytowań	h-indeks	Cytowania odzyskane	Zwiększenie h-indeksu
0000-0003-2351-...	200	7	3,5%	1
0000-0002-2529-...	595	12	0,7%	0
0000-0002-8650-...	22	2	4,5%	0
0000-0003-3817-...	40	3	2,5%	0
0000-0001-7741-...	250	6	8,8%	1
0000-0002-1049-...	21	3	0	0
0000-0002-6946-...	17	2	23,5%	1

Nr ORCID	Ostateczna liczba cytowań	h-indeks	Cytowania odzyskane	Zwiększenie h-indeksu
0000-0002-3720-...	131	6	4,6%	0
Średnio:			6%	0,4

Źródło: Opracowanie własne na podstawie wyników badań.

Maksymalne wartości określające liczbę błędów w obu bazach to ponad 70% dla Web of Science oraz ponad 23% dla Scopus. Korekta błędów w bazach przyniosła także zwiększenie h-indeksu maksymalnie o 2 lub 1, odpowiednio dla baz Web of Science oraz Scopus.

Ostatnia tabelka dotyczy wybranych polskich i zagranicznych czasopism naukowych. W tabelce przedstawiono opublikowane przez bazę Web of Science w czerwcu 2018 roku wartości parametru IF za rok 2017 oraz procentową wartość teoretycznego zwiększenia tego parametru w przypadku zgłoszenia wykrytych w bazie błędów (znajdzone błędy nie we wszystkich przypadkach zostały zgłoszone do bazy).

Tab. 3. Wartość parametru IF dla wybranych czasopism za rok 2017 w bazie Web of Science

Lp.	Nr ISSN	Czasopismo	IF 2017	Teoretyczny Wzrost IF 2017
1.	1335-8871	Measurement Science Review	1,345	3,6%
2.	1687-8132	Advances in Mechanical Engineering	0,848	23,1%
3.	1756-8277	International Journal of Spray and Combustion Dynamics	1,083	5,2%
4.	2072-666X	Micromachines	2,222	11,2%
5.	2073-4395	Agronomy-Basel	1,419	11,4%
6.	2073-8994	Symmetry-Basel	1,256	7,2%
7.	1687-725X	Journal of Sensors	2,057	33,1%
8.	1687-7969	Advances in Astronomy	1,209	11,6%
9.	1687-8086	Advances in Civil Engineering	0,827	44,2%
10.	1472-3646	Archaea-an International Microbiological Journal	1,756	41,7%
11.	2090-8016	Cardiology Research and Practice	1,143	60,9%
12.	1687-6075	Science and Technology of Nuclear Installations	0,939	22,2%
13.	1872-2105	Recent Patents on Nanotechnology	1,475	5,1%
14.	1570-1646	Current Proteomics	0,606	5,0%
15.	1573-4099	Current Computer-Aided Drug Design	0,770	19,2%
16.	1573-4056	Current Medical Imaging Reviews	0,299	-0,2%

Lp.	Nr ISSN	Czasopismo	IF 2017	Teoretyczny Wzrost IF 2017
17.	0137-5075	Archives of Acoustics	0,917	3,3%
18.	2083-4772	Archives of Environmental Protection	1,120	4,8%
19.	1733-3490	Archives of Metallurgy and Materials	0,625	6,3%
20.	0208-6425	Chemical and Process Engineering	0,892	6,0%
21.	0239-7528	Bulletin of the Polish Academy of Sciences-Technical Sciences	1,361	4,8%
22.	0860-8229	Metrology and Measurement Systems	1,523	2,5%
23.	1662-5218	Frontiers in Neurorobotics	2,606	17,4%
24.	2329-7662	3D Printing and Additive Manufacturing	2,304	7,7%
25.	0267-5730	International Journal of Technology Management	0,869	8,2%
Średnio:				15%

Źródło: Opracowanie własne na podstawie wyników badań.

Maksymalna wartość określająca błąd wyznaczenia parametru IF to ponad 60%. W przypadku bazy Scopus przeprowadzono podobne badania, jednak dla mniejszej liczby czasopism. Jak dotąd nie zaistniała sytuacja, w której wykryto błędy skutkujące widoczną zmianą parametru CS dla danego czasopisma. Wynika to ze znacznie lepszej jakości informacji w tej bazie, a także z faktu, że parametr CS wyznaczany jest w szerszej perspektywie czasowej, zarówno w odniesieniu do roczników publikacji cytujących, jak i cytowanych.

Jak dotąd największym osiągnięciem była poprawa parametru IF za 2018 r. dla polskiego czasopisma „Argumenta Oeconomica” wydawanego przez Uniwersytet Ekonomiczny we Wrocławiu. Dzięki odnalezieniu sześciu brakujących cytowań, IF dla tego czasopisma wzrósł z 0,245 do 0,347 (czyli o 42%), natomiast 5-year IF wzrósł z 0,301 do 0,366 (czyli o 22%).

Korekta błędów w bibliograficznych bazach danych

W przypadku odnalezienia publikacji, która powinna generować cytowanie, należy za pośrednictwem specjalnego formularza złożyć reklamację. W przypadku bazy Web of Science należy wybrać odnośnik *Data Correction* na samym dole strony internetowej lub *Suggest a correction* tuż nad listą cytowanej literatury dla danej publikacji. W przypadku bazy Scopus wybieramy odnośnik *Contact us* na samym dole strony in-

ternetowej, a następnie *How can I add missing citations?*. Wymagane jest wcześniejsze założenie własnego konta w tych bazach. Możliwe jest zgłaszanie kilku błędów na raz, przy wykorzystaniu odpowiedniego arkusza kalkulacyjnego, dostępnego w tych formularzach.

W bazie Web of Science typowe błędy poprawiane są w ciągu 3–5 dni, natomiast w bazie Scopus czas oczekiwania to 2–3 tygodnie. Możliwe jest także zgłoszenie reklamacji dotyczącej brakującej publikacji przez osobę, która jest pracownikiem danego czasopisma/wydawnictwa. W przypadku bazy Web of Science nie są przyjmowane takie zgłoszenia od osób niezwiązanych z danym wydawnictwem.

Wnioski

Przeprowadzone badania wykazały, że w bibliograficznych bazach danych istnieje duża liczba błędów w metadanych dotyczących konkretnej publikacji. Biorąc pod uwagę pojedynczych naukowców lub pojedyncze czasopisma, poziom błędów jest często nieznaczący i sięga kilku procent lub mniej. Natomiast zdarzają się przypadki, w których błędy przybierają dwucyfrowe wartości procentowe. Dotyczy to z oczywistych względów przede wszystkim autorów o małym dorobku publikacyjnym, rzadko cytowanym, a także czasopism wydających rocznie małą liczbę publikacji (np. kilkadziesiąt). Wówczas nawet pojedyncze cytowanie zauważalnie zmienia parametry bibliometryczne.

Na podstawie zaprezentowanych rezultatów badań można stwierdzić, że baza Scopus cechuje się zdecydowanie lepszą jakością informacji niż baza Web of Science. Co więcej, baza Scopus podaje parametr CS dla czasopism z dokładnością do jednego miejsca po przecinku i w dłuższej perspektywie czasowej (4 lata). Natomiast baza Web of Science (a właściwie Journal Citation Report) podaje parametr IF z dokładnością do trzech miejsc po przecinku, co sprawia, że często pojedyncze cytowanie zauważalnie zmienia jego wartość (w przypadku, gdy czasopismo wydaje w roku mniej niż 500 publikacji), co czyni tę bazę jeszcze mniej wiarygodną.

Podsumowując, należy wymienić jeszcze jeden mankament analizowanych baz danych. W niektórych przypadkach występują bardzo duże opóźnienia we wprowadzaniu do baz nowych publikacji (zwłaszcza referatów konferencyjnych – opóźnienia nierzadko sięgają ponad roku – zwłaszcza w przypadku Web of Science). Skutkiem tego może być zniżenie wartości parametrów bibliometrycznych (IF, CS) dla cytowanych przez takie publikacje czasopism, ponieważ są one wyznaczone raz

w roku. Z kolei do bazy Scopus wprowadzane są publikacje oznaczane jako *Article in Press*, które często przez długi okres nie posiadają listy cytowanych pozycji bibliograficznych, co może odnieść ten sam rezultat.

Dynamiczny rozwój technik informacyjnych, w tym zwłaszcza rozwój sztucznej inteligencji, pozwala mieć nadzieję na stałą poprawę jakości informacji w bibliograficznych bazach danych. Ważnym krokiem jest wymaganie posiadania indywidualnego numeru identyfikacyjnego naukowca (ORCID – ang. *Open Researcher and Contributor ID*) przez autorów publikacji. W ten sposób łatwo przypisać daną publikację do profilu autora. Kolejna ważna innowacja to stosowanie unikatowego numeru identyfikacyjnego dla każdej publikacji (DOI – ang. *Digital Object Identifier*). Niestety, nadal nie wszystkie publikacje zawierają te unikatowe identyfikatory.

Na koniec warto wspomnieć, że korzystnym podejściem jest wykorzystywanie bibliograficznych baz danych przy wyszukiwaniu literatury do aktualnie tworzonego artykułu. Dostępność wielu różnych filtrów i kryteriów wyszukiwania w tych bazach znacznie przyspiesza wyszukiwanie publikacji naukowych.

Bibliografia

- Babik W., *Ekologia informacji*, Kraków 2014.
- Clarivate, *Institute for Scientific Information: An introduction to global university rankings*, <https://clarivate.com/lp/an-introduction-to-global-university-rankings/> [dostęp: 20 listopada 2021 r.].
- Clarivate, *Web of Science Core Collection*, <https://clarivate.com/webofsciencigroup/solutions/web-of-science-core-collection/> [dostęp: 12 sierpnia 2021 r.].
- Elsevier, *Czym jest SciVal*, <https://www.elsevier.com/pl-pl/solutions/scival/scival-polska> [dostęp: 12 sierpnia 2021 r.].
- Elsevier, *Scopus*, https://www.elsevier.com/solutions/scopus?dgcid=RN_AGCM_Sourced_300005030 [dostęp: 12 sierpnia 2021 r.].
- Franceschini F., Maisano D., Mastrogiacomo L., *Empirical analysis and classification of database errors in Scopus and Web of Science*, „Journal of Informetrics” 2016, no. 10.4, s. 933–953.
- Gaca K., Psonka T., *Ocena jakości czasopism z wykorzystaniem parametrów bibliometrycznych w bazie Scopus*, „Społeczeństwo Informacyjne” 2017, z. 2, s. 15–28.
- Grudzień Ł., *Koncepcja oceny jakości informacji o procesach w systemach zarządzania*, [w:] *Innowacje w zarządzaniu i inżynierii produkcji*, red. R. Knosala, Opole 2012, s. 633–644.

- Kulczycki E., *Punktoza jako strategia w grze parametrycznej w Polsce*, „Nauka i Szkolnictwo Wyższe” 2017, nr 1(49), s. 63-78.
- Kulikowski K., Antipow E., *Niezamierzone konsekwencje punktozy jako wartości kulturowej polskiej społeczności akademickiej*, „Studia Socjologiczne” 2020, nr 3(238), s. 207-236.
- Mongeon P., Paul-Hus A., *The journal coverage of Web of Science and Scopus: a comparative analysis*, „Scientometrics”, 2016, no. 106, s. 213-228.
- Morawski R.Z., *Nauka w czasach biurokracji*, [w:] *Inżynier z duszą humanisty*, red. K. Baranowski, J. Sośnicka, Łódź, 2017, s. 321-348.
- Narodowe Centrum Nauki, *Raportowanie projektów NCN. Raport końcowy – wykaz publikacji*, <https://www.ncn.gov.pl/finansowanie-nauki/informacje-dla-realizujacych-projekty/raporty-ncn/koncowy/publikacje> [dostęp: 12 sierpnia 2021 r.].
- Rada Doskonałości Naukowej, *Wymagania dokumentacyjne wniosków w sprawie nadania tytułu profesora*, <https://www.rdn.gov.pl/postepowanie-profesorskie.wymagania-dokumentacyjne-wnioskow-w-sprawie-nadanai-tytulu-profesora.html> [dostęp: 12 sierpnia 2021 r.].
- Ruževičius J., Gedminaitė A., *Business information quality and its assessment*, „Engineering Economics” 2007, no. 52.2, s. 18-25.
- Singh V.K. et al., *The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis*, „Scientometrics” 2021, no. 126, s. 5113-5142.
- Stefanowicz B., *Informacja*, Warszawa 2004.
- Szczerbowski T., *Punktoza jako słowo ostatnich lat*, „Poradnik Językowy” 2017, z. 7, s. 80-87.
- Vera-Baceta M.A., Thelwall M., Kousha K., *Web of Science and Scopus language coverage*, „Scientometrics” 2019, no. 121, s. 1803-1813.
- Włoskovicz W., *Nowomowa pseudometanaukowa? O języku polskiej polityki naukowej*, „Poradnik Językowy” 2019, z. 10, s. 35-47.
- Żywczoł A., *Współczesny uniwersytet – w pułapce racjonalności eksplikatywnej*, „Rocznik Lubuski” 2018, t. 44, cz. 2, s. 93-106.