

Ilościowa analiza wykorzystania oprogramowania w badaniach bibliometrycznych

Adam Jachimczyk

ORCID 0000-0003-2917-6926

Katedra Bibliografii i Dokumentacji

Wydział Dziennikarstwa, Informacji i Bibliologii

Uniwersytet Warszawski

Abstrakt

Cel/Teza: Celem analizy było prześledzenie częstości wykorzystania oprogramowania bibliometrycznego oraz aplikacji do analizy sieci społecznych przez badaczy publikujących teksty o tematyce bibliometrycznej.

Koncepcja/Metody badań: Artykuły wraz z danymi bibliograficznymi zostały pobrane z serwisu Public Library of Science (PLOS). Do zidentyfikowania publikacji wykorzystano pakiet rplos. Nazwy aplikacji stosowanych w analizach bibliometrycznych ustalono na podstawie literatury przedmiotu oraz witryn internetowych poświęconych tej problematyce. Ogółem w analizie uwzględniono 52 aplikacje, w tym 38 bibliometrycznych i 14 programów do analizy sieci społecznych.

Wyniki i wnioski: Łącznie znaleziono 144 artykuły, w których było wymienione przynajmniej jedno oprogramowanie bibliometryczne lub do analizy sieci społecznych. W publikacjach wymieniono 57.69% aplikacji spośród 52, które stały się przedmiotem analizy. Badacze wykorzystali przynajmniej raz 52.63% aplikacji bibliometrycznych oraz 71.43% programów do analizy sieci społecznych. Wśród oprogramowania wyraźnie dominują dwa programy: Gephi i VOSviewer, z których każdy był wskazywany przez badaczy w ponad 20% artykułów.

Oryginalność/Wartość poznawcza: Analiza pokazuje znaczenie określonego oprogramowania w analizach bibliometrycznych. Identyfikuje najczęściej wykorzystywane oprogramowanie oraz ewolucję jego wykorzystania w ostatniej dekadzie.

Słowa kluczowe

Analiza sieci społecznych. Bibliometria. Oprogramowanie. Wykorzystanie oprogramowania.

Otrzymany: 23 października 2021. Zrecenzowany: 12 listopada 2021. Poprawiony: 17 grudnia 2021. Zaakceptowany: 23 grudnia 2021.

1. Wstęp

W piśmiennictwie naukowym generalnie widoczny jest wzrost liczby badań bibliometrycznych. Na przykład w porównaniu z dekadą 2001–2010, w latach 2011–2020 liczba artykułów zarejestrowanych w bazie Scopus, które zaindeksowano terminem „bibliometrics” zwiększyła się blisko czterokrotnie, z 3639 do 13931 (stan z 15 grudnia 2021 r.). Dane te potwierdzają obserwowany w wielu krajach wzrost znaczenia bibliometrii (czy szerzej – naukometrii) w badaniu i ocenie dorobku naukowego badaczy i jednostek naukowych oraz

w badaniu kierunków rozwoju współczesnej nauki. Nie bez znaczenia jest także szeroka dostępność różnego typu źródeł informacji, które dostarczają materiału do wielowątkowych badań bibliometrycznych. Większość analiz opiera się na bibliograficznych bazach danych, takich jak: Web of Science, Scopus, czy PubMed, ale badacze niejednokrotnie korelują dane ze wspomnianych źródeł z innymi zasobami: danymi statystycznymi, rejestrami medycznymi, danymi z krajowych systemów informacji naukowej, czy sięgają po dane z mediów społecznościowych. Większość źródeł informacji udostępnia ponadto dane w formie, która ułatwia ich ponowne wykorzystanie dzięki ogólnie dostępnemu oprogramowaniu do selekcji, analizy i wizualizacji informacji.

Właśnie odpowiednie oprogramowanie znacząco przyspiesza i wzbogaca badania bibliometryczne. Duża liczba danych wymaga zastosowania odpowiednich narzędzi ułatwiających gromadzenie materiału badawczego, jego analizę i prezentację. Arsenal aplikacji stosowanych przez naukowców jest bardzo rozbudowany. Obejmuje on m.in. oprogramowanie do analiz statystycznych, systemy zarządzania bazami danych, aplikacje do zarządzania metadanymi bibliograficznymi, procesory tekstu, arkusze kalkulacyjne, specjalistyczne oprogramowanie do analizy danych jakościowych, czy eksploracji tekstu (ang. *text mining*). Specjalną grupę stanowią aplikacje przeznaczone wyłącznie do analiz bibliometrycznych i analizy sieci społecznych.

Nazwę *oprogramowanie (aplikacje) bibliometryczne* odnosi się do oprogramowania, które umożliwia obliczanie wskaźników bibliometrycznych, wykorzystując informacje pochodzące z bibliograficznych baz danych (Todeschini & Baccini, 2016). Tego rodzaju oprogramowanie można podzielić na dwie kategorie:

- (1) oprogramowanie przeznaczone do analizy bibliometrycznej dorobku naukowego pojedynczych badaczy, grup badawczych, jednostek naukowych, który mierzony jest liczbą publikacji i/lub liczbą cytowań (np. CRExplorer, Publish or Perish);
- (2) oprogramowanie do analizy relacji zachodzących w nauce. Analiza sieci społecznych (ang. *social network analysis*) znajduje szerokie zastosowanie w tym zakresie, pozwalając na badanie formalnych i nieformalnych relacji, które ułatwiają lub blokują przepływ wiedzy między podmiotami naukowymi (Serrat, 2017). Aplikacje bibliometryczne dostarczają szeregu narzędzi do badania kierunków rozwoju nauki, analizując sieci powiązań między publikacjami naukowymi, ich autorami, czy podejmowanymi przez nich tematami badawczymi (np. VOSviewer, bibliometrix, CiteSpace) (Moral-Muñoz et al., 2020). Są one często zastępowane lub uzupełniane przez aplikacje do analizy sieci społecznych, które pierwotnie nie były projektowane pod kątem analizy danych bibliograficznych (np. Gephi, UCINET, Pajek).

Niniejszy artykuł ma wypełnić pewną lukę w piśmiennictwie poświęconym bibliometrii, w którym nie brakuje opracowań na temat źródeł informacji (Archambault et al., 2009; Meho & Yang, 2007), wskaźników bibliometrycznych (Cooper, 2015; Roldan-Valadez et al., 2019), czy opisów oprogramowania wykorzystywanego w badaniach (por. Cobo et al., 2011; Moral-Muñoz et al., 2020). Stosunkowo niewiele opracowań dotyczy natomiast samego warsztatu współczesnych badań bibliometrycznych, a w szczególności wykorzystania w nich specjalistycznego oprogramowania bibliometrycznego, które obecnie jest reprezentowane przez przynajmniej kilkadziesiąt aplikacji.

2. Przegląd literatury

Badacze analizujący procesy zachodzące w nauce coraz częściej zwracają uwagę na znaczenie oprogramowania w badaniach naukowych. Silnie akcentują np. potrzebę właściwego cytowania wykorzystywanych programów. W tym kierunku idą prace grupy naukowców skupionych wokół FORCE11, którzy podkreślają, że oprogramowanie powinno być traktowane w wykazie piśmiennictwa i w cytowaniach na takich samych zasadach, jak artykuły, książki czy dane badawcze (Smith et al., 2016). Brak standardów dotyczących cytowania oprogramowania w literaturze utrudnia jego indeksowanie, wyszukiwanie i docenienie roli, jaką odgrywa we współczesnej nauce (Du et al., n.d.). Precyzyjne cytowanie ułatwia bowiem dostęp do zbiorów danych, które niejednokrotnie mogą być nieczytelne bez ściśle określonego oprogramowania (Starr et al., 2015). Jest także kluczowe dla celów związanych z replikacją badania (Howison & Bullard, 2016a). Wymienienie w artykule nazwy wykorzystanego oprogramowania niesie ze sobą także niebagatelny walor edukacyjny, kierując uwagę innych badaczy, zwłaszcza młodszych, na istnienie narzędzi przydatnych w analizach naukowych (Pan et al., 2019). Taki cel przyświecał np. analizie przeprowadzonej przez A.E. Dembe, J.S. Partridge i L.C. Geist (Dembe et al., 2011), którzy badali wykorzystanie pakietów do analizy statystycznej przez autorów publikujących artykuły poświęcone służbie zdrowia.

Duża liczba badaczy wykorzystujących określone oprogramowanie świadczy pośrednio o jego jakości oraz stwarza możliwości uzyskania pomocy w zakresie jego wykorzystania (Howison et al., 2015). O znaczeniu, jakie niesie wskazanie oprogramowania w publikacji, świadczą wyniki badania sondażowego przeprowadzonego przez Joppę z zespołem, które pokazało, że blisko 30% ankietowanych przez nich respondentów przy wyborze określonego oprogramowania kierowało się tym, iż jego użycie zostało potwierdzone w recenzowanych czasopismach (Joppa et al., 2013).

Oprogramowanie jako istotna część infrastruktury badawczej dopiero od stosunkowo niedawna stało się przedmiotem refleksji naukowej. Studia dotyczące oprogramowania koncentrują się głównie na zagadnieniach związanych z jego rozwojem, archiwizacją i cytowaniem w publikacjach. Nieco rzadziej podejmowane są analizy dotyczące użytkowania oprogramowania oraz jego wpływu na badania naukowe (Mayernik et al., 2017).

Badaczy interesuje np. kwestia wykazywania w artykułach oprogramowania oraz jego poprawnego cytowania. Pan z zespołem, odwołując się do analizy korpusu tekstów opublikowanych w PLoS ONE w 2014 r., wykazali, że autorzy artykułów 1868 razy odnotowali wykorzystanie w badaniu oprogramowania SPSS – pakietu do analizy statystycznej, ale tylko 42 razy oprogramowanie to zostało wymienione w formalnym cytowaniu (Pan et al., 2015).

Statystyki dotyczące wskazywania w publikacjach naukowych oprogramowania wykorzystanego w badaniach różnią się w zależności od tego, jaki obszar nauki był badany i na jakim korpusie tekstów oparto badanie. Pan z zespołem, na podstawie analizy blisko 572 publikacji z zakresu nauki o informacji i bibliotekoznawstwa, ustalili, że nazwa oprogramowania użytego w badaniu została wymieniona tylko w 27% publikacji (Pan et al., 2019). Du z zespołem, badając korpus 4971 tekstów medycznych i ekonomicznych, ustalili, że 71% artykułów nie zawiera żadnych informacji o wykorzystanym oprogramowaniu (Du et al., 2021). Z kolei Howison i Bullard zidentyfikowali wzmianki o oprogramowaniu w 65% spośród 90 artykułów reprezentujących obszar nauk biologicznych (Howison & Bullard, 2016b).

Cytowany już Dembe z zespołem nazwy wykorzystanego oprogramowania wykrył w 65% artykułów opublikowanych w amerykańskich czasopismach poświęconych służbie zdrowia (Dembe et al., 2011). Stosunkowo wysoki odsetek, wynoszący blisko 80%, wykazywania nazw oprogramowania w artykułach naukowych zidentyfikował Pan z zespołem w grupie 9548 artykułów z różnych dziedzin wiedzy opublikowanych w PLoS ONE. Najczęściej, w 86% przypadków, wykazywano oprogramowanie w publikacjach z zakresu nauk rolniczych, najrzadziej (61%) w artykułach z dziedziny matematyki (Pan et al., 2016).

Cytowani badacze uważają, że autorzy niejednokrotnie nie widzą potrzeby wskazywania na wykorzystanie popularnego oprogramowania (np. Microsoft Excel, Microsoft Word), gdyż nie wpływa ono na rezultaty badania. Część autorów w ogóle nie wymienia programów użytych w badaniu w artykułach prezentujących jego wyniki (Dembe et al., 2011; Pan et al., 2019). Można przypuszczać, że ta sytuacja będzie się z upływem czasu zmieniała, gdyż coraz więcej redakcji czasopism, jak np. PLoS ONE, w trosce o jakość publikowanych badań, wymaga lub przynajmniej zachęca autorów do udostępniania wykorzystanego oprogramowania lub algorytmu (PLoS ONE, n.d.).

Niektóre badania przedstawione w piśmiennictwie mogą być przydatne np. w analizie wzorców wykorzystania oprogramowania w nauce (cytowania, udostępnianie) czy przyczynić się do rozwoju usług związanych z oprogramowaniem dla badaczy (Du et al., n.d.; Schindler et al., 2020). Kluczowe znaczenie odgrywa tutaj opracowanie korpusów nazw oprogramowania wymienianego przez autorów w artykułach naukowych. Przykładem takiego zestawu danych jest SoftwareKG, w którym badacze zgromadzili dane o oprogramowaniu z 51 tys. artykułów naukowych reprezentujących nauki społeczne, opublikowanych w serwisie PLoS (Schindler et al., 2020).

Badacze kilkakrotnie podejmowali także kwestię oceny oraz wykorzystania oprogramowania bibliometrycznego. Pan z zespołem zbadał np. wykorzystanie oraz cytowanie trzech aplikacji: CiteSpace, HistCite oraz VOSviewer w 481 artykułach opublikowanych w latach 2002–2017. Z ich analizy wynika, że do wspomnianych narzędzi odwoływali się autorzy reprezentujący obszar nauki o informacji i bibliotekoznawstwa oraz nauk informatycznych. Najczęściej wykorzystywanymi aplikacjami były CiteSpace i VOSviewer (Pan et al., 2018). Pośrednio stopień wykorzystania oprogramowania bibliometrycznego można oszacować na podstawie korpusu SoftwareKG. Na przykład, VOSviewer i HistCite zostały wymienione 11 razy. Znacznie częściej wymieniano oprogramowanie do analizy sieci społecznych. Gephi oraz Pajek łącznie zostały wymienione ponad 100 razy (Schindler et al., 2019).

W literaturze naukowej nie brakuje także charakterystyk oprogramowania bibliometrycznego (Bales et al., 2020; Cobo et al., 2011; Moral-Munoz et al., 2019; Moral-Muñoz et al., 2020; Pradhan, 2016). Moral-Muñoz z zespołem (Moral-Muñoz et al., 2020) w konkluzji swojej wielostronnej analizy uznali pakiet bibliometrix za oprogramowanie o najbardziej rozbudowanej liczbie funkcji, a VOSviewer został najlepiej oceniony za możliwości wizualizacji danych. Cobo z zespołem (Cobo et al., 2011) podkreślili, że kompletna analiza bibliometryczna powinna zostać wykonana z wykorzystaniem kilku aplikacji, pokazując w ten sposób materiał z różnych perspektyw. Oprogramowanie różni się bowiem pod względem dostępnych funkcji, choćby w zakresie wstępnej obróbki danych, czy sposobów ich wizualizacji.

Uwagę środowiska naukowego przykuwa także rosnąca popularność darmowego pakietu R do analizy statystycznej i programowanego przy użyciu języka o tej samej nazwie,

stosowanego przez przedstawicieli różnych dyscyplin naukowych (Gentleman et al., 2004). Li i Yan (Li & Yan, 2018) ustalili, na podstawie analizy ponad 13 600 publikacji z serwisu PLoS, że w prawie 55% stosowany był jeden z pakietów R. Stanowi on znaczącą alternatywę dla komercyjnego oprogramowania stosowanego przez badaczy (np. SPSS firmy IBM, Stata firmy StataCorp). Już w marcu 2011 r. język R znajdował się w grupie 30 najpopularniejszych języków programowania (German et al., 2013). Swoją popularność zawdzięcza pakietom – rozszerzeniom, pracującym w środowisku R (Decan et al., 2015), które tworzą jego użytkownicy. W oficjalnym repozytorium pakietów Comprehensive R Archive Network (CRAN, n.d.) są dostępne także narzędzia przeznaczone do analiz bibliometrycznych, w tym m.in. zyskujący na znaczeniu pakiet bibliometrix (Aria & Cuccurullo, 2017).

3. Cel badania

Celem analizy było prześledzenie częstości wykorzystania oprogramowania bibliometrycznego oraz aplikacji do analizy sieci społecznych przez badaczy publikujących teksty o tematyce bibliometrycznej.

Starano się uzyskać odpowiedź na następujące pytania:

- (1) Jakie oprogramowanie do analizy bibliometrycznej oraz do analizy sieci społecznych wykazywali badacze w opublikowanych artykułach oraz jak często odwoływali się do poszczególnych aplikacji?
- (2) W jakim odsetku publikacji badacze wykazywali stosowanie innego rodzaju oprogramowania?
- (3) W jakim odsetku publikacji badacze odwoływali się do więcej niż jednej aplikacji w opublikowanych artykułach?

Ilościowa analiza pokazuje znaczenie określonego oprogramowania w analizach bibliometrycznych oraz ewolucję częstości jego wykorzystania w latach 2009–2021. Wyniki tej analizy mogą posłużyć jako wskazówka dla innych badaczy w zakresie wyboru odpowiedniego narzędzia do analiz bibliometrycznych.

4. Metody

W artykule przeprowadzono ilościową analizę wykorzystania aplikacji bibliometrycznych w artykułach opublikowanych w Public Library of Science (PLoS, n.d.). Serwis zapewnia otwarty dostęp do pełnych tekstów ponad 200 tys. artykułów reprezentujących różne obszary nauki, w przeważającej mierze – nauki biologiczne i medyczne.

Do zidentyfikowania publikacji wykorzystano pakiet rplos (Chamberlain et al., 2021) uruchamiany w środowisku R (R Foundation, 2020). Pakiet, dzięki interfejsowi API udostępnionemu przez wydawcę PLoS, umożliwia przeszukiwanie bazy danych oraz pobieranie opublikowanych artykułów.

W celu przeszukania zasobu PLoS zastosowano funkcję searchplos:

```
searchplos(q=„bibliometrics”, fq=’body:„[nazwa_oprogramowania]”).
```

W ten sposób sformułowana kwerenda wyszukuje artykuły, w których tekście występuje słowo kluczowe „bibliometrics”. Argument fq=„body” pełni funkcję filtra ograniczając pulę odszukanych publikacji tylko do tych, w których treści (z wyjątkiem abstraktu i literatury przedmiotu) występuje nazwa określonego oprogramowania.

Zdecydowano się na użycie w kwerendzie słowa kluczowego „bibliometrics”, które ma szeroki zakres znaczeniowy, pozwalający zidentyfikować publikacje dotyczące bibliometrii. Ponieważ nie zawsze artykuły omawiające analizy bibliometryczne są zaindeksowane terminem „bibliometrics”, w przeszukiwaniu zasobu PLoS został też użyty warunek wystąpienia tego wyrażenia w treści dokumentu.

Nazwy aplikacji bibliometrycznych ustalono na podstawie literatury przedmiotu oraz witryn internetowych poświęconych tej problematyce. Ogółem zidentyfikowano 52 aplikacje, w tym 38 programów do analiz bibliometrycznych i 14 programów do analizy sieci społecznych.

Tab. 1. Oprogramowanie bibliometryczne

Lp.	Nazwa aplikacja	Rodzaj aplikacji	Źródło
1	2	3	4
1.	Ambalytics	B	(Kammerer et al., 2021)
2.	Bibexcel (Bibexcel)	B	(Moral-Muñoz et al., 2020)
3.	Bibliographic Item Co-Occurrence Matrix Builder	B	(Xu et al., 2013)
4.	Bibliographic Item Co-Occurrence Mining System	B	(Li et al., 2016)
5.	bibliometrix	B	(Moral-Muñoz et al., 2020)
6.	Bibliothécaire	B	(Bouchard et al., 2015)
7.	BiblioTools (interfejs webowy BiblioMaps)	B	(Moral-Muñoz et al., 2020)
8.	Citan	B	(Moral-Muñoz et al., 2020)
9.	CiteSpace	B	(Moral-Muñoz et al., 2020)
10.	CitNetExplorer	B	(Moral-Muñoz et al., 2020)
11.	CoCites	B	(Kammerer et al., 2021)
12.	Connected Papers	B	(Kammerer et al., 2021)
13.	CoPalRed	B	(Pradhan, 2016)
14.	Cortext	B	(Elie et al., 2021)
15.	CRExplorer (Cited Reference Explorer)	B	(Moral-Muñoz et al., 2020)
16.	Headstart	B	(Hamid Jamali, n.d.)
17.	HistCite	B	(Pradhan, 2016)
18.	iCite	B	(Hutchins et al., 2019)
19.	Inciteful	B	(Kammerer et al., 2021)

1	2	3	4
20.	IN-SPIRE	B	(Pradhan, 2016)
21.	Loet Leydesdorff software	B	(Pradhan, 2016)
22.	Metaknowledge	B	(Moral-Muñoz et al., 2020)
23.	Meva (MEDLINE Evaluator)	B	(Tenner et al., 2003)
24.	Network Workbench	B	(Pradhan, 2016)
25.	Publish or Perish	B	(Moral-Muñoz et al., 2020)
26.	PubMed PubReminer	B	(Slater, 2012)
27.	SAINT (Science Assessment Integrated Network Toolkit)	B	(Hamid Jamali, n.d.)
28.	Scholarometer	B	(Kaur et al., 2012; 2012b)
29.	Science of Science Tool (Sci2)	B	(Moral-Muñoz et al., 2020)
30.	ScientoPY (ScientoPyUI)	B	(Moral-Muñoz et al., 2020)
31.	scientoText	B	(Moral-Muñoz et al., 2020)
32.	SciMAT (Science Mapping Anaylsis Tool)	B	(Moral-Muñoz et al., 2020)
33.	Sitkis	B	(Pradhan, 2016)
34.	Tethne	B	(Kammerer et al., 2021)
35.	VantagePoint	B	(Pradhan, 2016)
36.	VOSviewer (VOS-viewer)	B	(Moral-Muñoz et al., 2020)
37.	Webometric Analyst	B	(Hamid Jamali, n.d.)
38.	WoS2Pajek	B	(Pajek, 2007)
39.	Cytoscape	N	(van Eck & Waltman, 2014)
40.	Gephi	N	(Majeed et al., 2020)
41.	GraphStream	N	(Social network, 2021)
42.	Graph-tool	N	(Social network, 2021)
43.	Graphviz	N	(Social network, 2021)
44.	iGraph library	N	(Social network, 2021)
45.	Java Universal Network/Graph (JUNG) Framework	N	(Social network, 2021)
46.	MuxViz	N	(Majeed et al., 2020)
47.	NetDraw	N	(Cronin, 2015)
48.	NetMiner	N	(Social network, 2021)
49.	NetworkX	N	(Social network, 2021)
50.	NodeXL	N	(Majeed et al., 2020)
51.	Pajek	N	(Majeed et al., 2020)
52.	UCINET	N	(Majeed et al., 2020)

B – oprogramowanie do analizy bibliometrycznej, N – oprogramowanie do analizy sieci społecznych

W czasie wyszukiwania w PLoS zastosowano także zidentyfikowane w literaturze przedmiotu różne warianty pisowni (również błędne) nazw oprogramowania stosowanego przez badaczy (BibExcel był zapisywany także jako BibExel, VOSviewer jako VOS-viewer). Nazwy aplikacji przedstawia tabela 1. W nawiasie umieszczono wariant nazwy programu.

Identyfikując program, nie rejestrowano informacji o jego wersji, tylko sam fakt jego wykorzystania. Niewątpliwie kolejne aktualizacje, poprawiające błędy i dodające nowe funkcje do aplikacji, mogą do pewnego stopnia wpływać na częstość jej wykorzystania, ale to zagadnienie wymaga dalszych pogłębionych badań.

W analizie pominięto publikacje, których autorzy wykorzystali wskaźniki bibliometryczne generowane dynamicznie w bibliograficznych bazach danych, takich jak Scopus czy Web of Science. Celem było bowiem zidentyfikowanie autonomicznych aplikacji, niezintegrowanych z żadną bazą danych, przetwarzających dane pochodzące z różnych źródeł.

Po zidentyfikowaniu grupy publikacji przejrzano je i usunięto te, w których wymieniono dane oprogramowanie, ale nie zostało ono wykorzystane do przeprowadzenia analizy (Hutchins et al., 2019; Kaur et al., 2012a).

Obliczenia statystyczne oraz wykres zostały wykonane w arkuszu kalkulacyjnym Microsoft Excel (Microsoft, n.d.). Wizualizację współwystępowania aplikacji opracowano w aplikacji VOSviewer, wersja 1.6.17 (VOSviewer, 2021).

Dane zostały zebrane i opracowane między 31 sierpnia a 10 września 2021 r.

5. Ograniczenia badania

Przyjęta metodologia nie identyfikuje wszystkich programów wykorzystywanych w analizach bibliometrycznych i sieci społecznych w nauce, np. takiego, które zostało specjalnie opracowane na potrzeby danego badania. Naukowcy tworzą takie aplikacje, jeśli istniejące oprogramowanie nie zaspokaja ich potrzeb (Cobo et al., 2011) albo wtedy, gdy brak darmowych programów o określonych funkcjonalnościach.

Próbę badawczą ogranicza też wspomniana wcześniej praktyka pomijania przez autorów artykułów wskazania wykorzystywanych aplikacji. Jest to dość powszechny problem, który niejednokrotnie uniemożliwia replikację badania i zweryfikowanie poprawności uzyskanych rezultatów. Analizę utrudnia również niejednorodność, z jaką zapisywane są nazwy stosowanych aplikacji. Na przykład w korpusie SoftwareKG badacze odnotowali użycie kilku różnych wersji nazwy pakietu SPSS: Statistical Package for the Social Sciences, IBM SPSS statistics, Statistical Package for Social Sciences, IBM SPSS (Schindler et al., 2020). Nie zawsze możliwe jest więc ustalenie wszystkich istniejących wariantów nazwy określonej aplikacji.

Badanie nie identyfikuje także pełnej skali wykorzystania innego rodzaju oprogramowania, które niewątpliwie jest wykorzystywane w analizach bibliometrycznych. Badacze powszechnie wykorzystują bowiem np. pakiety do analiz statystycznych (np. SPSS, Stata, pakiet R), arkusz kalkulacyjny (Microsoft Excel), narzędzia do zarządzania danymi bibliograficznymi (EndNote, ProCite, RefWorks), systemy zarządzania bazami danych (Microsoft Access, MySQL) czy narzędzia do analizy i wizualizacji danych (Tableau Public, QlikView).

Analiza objęła pierwsze osiem miesięcy 2021 r., więc prawdopodobnie niedoszacowana będzie liczba aplikacji wykorzystywanych przez autorów tekstów opublikowanych od

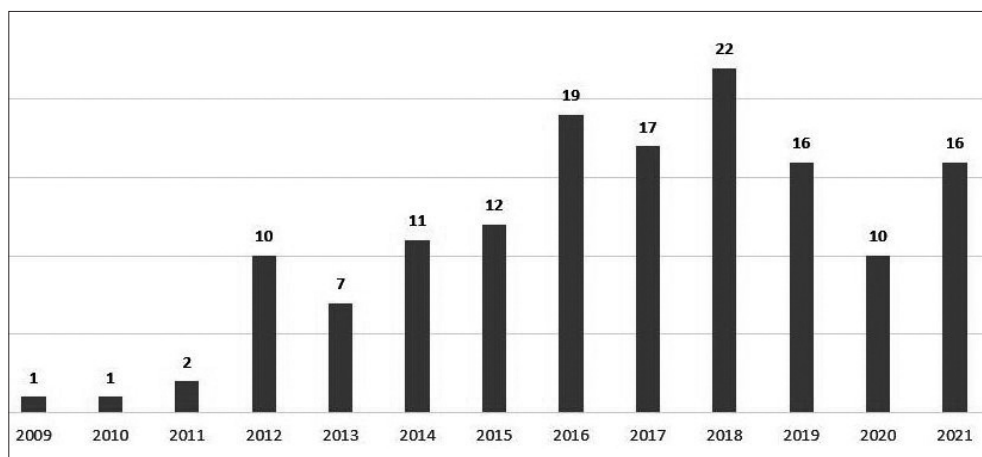
września do grudnia 2021 r. Publikacje zawarte w PLoS wydawane są od 2003 r., kiedy serwis rozpoczął działalność.

Trzeba też zaznaczyć, że nie we wszystkich publikacjach, prezentujących ilościowe analizy piśmiennictwa przeprowadzone za pomocą metod bibliometrycznych, analizy te są przez samych autorów identyfikowane jako badania czy analizy bibliometryczne, co skutkuje nieużywaniem terminu „bibliometrics” ani w tekście artykułu, ani w jego metadanych.

6. Wyniki analizy

6.1. Wykorzystanie oprogramowania do analizy bibliometrycznej i do analizy sieci społecznych

Łącznie odszukano 144 artykuły, w których wskazano wykorzystanie przynajmniej jednego oprogramowania bibliometrycznego. Reprezentują one 15.43% wszystkich publikacji udostępnionych w serwisie PLoS (łącznie 933), w których występuje słowo kluczowe „bibliometrics”. Liczbowy rozkład publikacji wg roku wydania przedstawia wykres (Rys. 1).



Rys. 1. Liczba artykułów, w których wskazano stosowanie oprogramowania bibliometrycznego z podziałem wg roku publikacji

W publikacjach wymieniono 57.69% aplikacji spośród 52 wcześniej zidentyfikowanych, które stały się przedmiotem analizy; 33.33% programów wskazano tylko raz, więcej niż 10 razy wymieniono 26.67% programów. Listę aplikacji przedstawia tabela 2.

Badacze wykorzystali przynajmniej raz 52.63% aplikacji do analiz bibliometrycznych oraz 71.43% programów do analizy sieci społecznych. Łącznie wszystkie aplikacje wymieniono w artykułach 210 razy, w tym w ponad połowie przypadków (53.81%) odwoływano się do oprogramowania przeznaczonego do analizy sieci społecznych. Wyraźnie dominują dwa programy: Gephi i VOSviewer, z których każdy był wykazywany przez badaczy w ponad 20% artykułów.

Tab. 2. Oprogramowanie wykorzystywane w badaniach bibliometrycznych opublikowanych w serwisie PLoS w latach 2009–2021

Lp.	Rodzaj aplikacji	Nazwa aplikacji	Liczba wystąpień	% (n=144)
1.	N	Gephi	36	25.00
2.	B	VOSviewer	31	21.53
3.	N	Pajek	20	13.89
4.	N	iGraph	17	11.81
5.	N	UCINET	14	9.72
6.	B	CiteSpace	12	8.33
7.	B	VantagePoint	11	7.64
8.	N	NetDraw	9	6.25
9.	N	Cytoscape	8	5.56
10.	B	iCite	7	4.86
11.	B	Sci2 Tool	7	4.86
12.	B	BibExcel	6	4.17
13.	B	bibliometrix	4	2.78
14.	B	CorText	3	2.08
15.	N	NetworkX	3	2.08
16.	N	NodeXL	3	2.08
17.	B	Publish or Perish	3	2.08
18.	B	CitNetExplorer	2	1.39
19.	N	NetMiner	2	1.39
20.	B	Network Workbench	2	1.39
21.	B	BiblioTools	1	0.69
22.	B	BICOMS	1	0.69
23.	N	Graph-tool	1	0.69
24.	B	IN-SPIRE	1	0.69
25.	B	PubReminer	1	0.69
26.	B	SAINT	1	0.69
27.	B	Scimat	1	0.69
28.	B	Tethne	1	0.69
29.	B	Webometric Analyst	1	0.69
30.	B	Wos2Pajek	1	0.69
Razem			210	

B – oprogramowanie do analizy bibliometrycznej, N – oprogramowanie do analizy sieci społecznych

Częstość wykorzystania oprogramowania bibliometrycznego zbadano w trzech okresach: 2009–2013, 2014–2017 i 2018–2021. Dane zostały przedstawione w tabeli 3.

W latach 2009–2013 łącznie wykazano w publikacjach wykorzystanie 16 aplikacji. Najczęściej wskazywano na oprogramowanie do analizy sieci społecznych, które wymieniono w ponad 60% artykułów. Okres ten cechuje stosunkowo niska liczba wykorzystywanego oprogramowania, ale niewielka jest także liczba zidentyfikowanych artykułów.

Tab. 3. Wykorzystanie oprogramowania bibliometrycznego i do analizy sieci społecznych w latach 2009–2021

Lp.	Rodzaj aplikacji	Program	2009–2013	2014–2017	2018–2021
1.	B	BibExcel	1	3	2
2.	B	bibliometrix			4
3.	B	BiblioTools			1
4.	B	BICOMS		1	
5.	B	CiteSpace	2	1	9
6.	B	CitNetExplorer		2	
7.	B	CorText	1	1	1
8.	N	Cytoscape	2	4	2
9.	N	Gephi	2	17	17
10.	N	Graph-tool		1	
11.	B	iCite		3	4
12.	N	iGraph	2	7	8
13.	B	IN-SPIRE		1	
14.	N	NetDraw	5	3	1
15.	N	NetMiner			2
16.	B	Network Workbench		2	
17.	N	NetworkX	1	1	1
18.	N	NodeXL	1	1	1
19.	N	Pajek	3	11	6
20.	B	Publish or Perish	1	2	
21.	B	PubReminer	1		
22.	B	SAINT		1	
23.	B	Sci2 Tool	3	3	1
24.	B	Scimat			1
25.	B	Tethne			1
26.	N	UCINET	5	6	3
27.	B	VantagePoint	3	6	2
28.	B	VOSviewer	1	11	19
29.	B	Webometric Analyst		1	
30.	B	Wos2Pajek		1	
Razem w poszczególnych okresach			34	90	86

B – oprogramowanie do analizy bibliometrycznej, N – oprogramowanie do analizy sieci społecznych

Między 2014 a 2017 r. dziewięć aplikacji do analizy sieciowej zostało wymienionych 51 razy. Tylko dwie aplikacje – Gephi i Pajek zostały wymienione łącznie w prawie 50% analizowanych publikacji. W tym okresie wzrosła również liczba wykorzystywanego oprogramowania, z 16 do 24.

Między 2018 a 2021 r. w nieco większej liczbie publikacji, w porównaniu z poprzednim okresem, spadła liczba wskazań oprogramowania, jak również liczba wykorzystywanych aplikacji. W tym okresie dziewięć programów do analizy sieciowej zostało wymienionych 41 razy. Jest to jedyny okres, kiedy oprogramowanie do analiz bibliometrycznych było nieco częściej wskazywane przez badaczy niż programy do analizy sieci społecznych. Na taki stan rzeczy niewątpliwym wpływ miało oprogramowanie VOSviewer i CiteSpace, wskazane w ponad 40% analizowanych publikacji. W tym okresie zwraca uwagę częste wykorzystywanie oprogramowania CiteSpace oraz pojawienie się aplikacji bibliometrix, zyskującej coraz większą popularność. O ponad połowę spadło wykorzystanie oprogramowania Pajek oraz VantagePoint. Na mniej więcej tym samym poziomie kształtowało się wykorzystanie programu BibExcel.

Z danych zawartych w tabeli 3 wyłania się także grupa aplikacji, które były stosowane w całym analizowanym okresie. Jest ich 14 (46.67%), w tym osiem programów do analizy sieci społecznych. W stosunku do pierwszego analizowanego okresu, w drugim pojawiło się dziewięć nowych aplikacji, ale tylko jedna z nich została wymieniona w trzecim okresie. W latach 2018–2021 liczba nowych aplikacji, niewymienionych we wcześniejszych latach, zmalała do pięciu, jednocześnie spadła w tym czasie liczba wykorzystywanych programów.

6.2. Wykorzystanie innego rodzaju oprogramowania

W mniej niż połowie artykułów (47.22%) autorzy, poza omawianymi aplikacjami, wskazali także szereg innych programów wykorzystywanych w analizach. Ogółem zidentyfikowano 69 aplikacji. Ponad 80% (84.06%) aplikacji zostało wymienionych tylko raz. Około 19% stanowi oprogramowanie opracowane przez autorów artykułów. Pięć z tych programów napisano specjalnie na potrzeby bibliometrycznej analizy. W ośmiu przypadkach autorzy wykorzystali język programowania Python, dwa razy Java, raz C++. W dwóch przypadkach nie sprecyzowano języka oprogramowania.

Autorzy nie zawsze publikują kod źródłowy samodzielnie stworzonego oprogramowania. Wśród 13 aplikacji specjalnie opracowanych na potrzeby określonej analizy, w Internecie zostały udostępnione tylko cztery. W jednym przypadku autorzy zaznaczyli, że aplikacja będzie dostępna na życzenie zainteresowanych osób.

Wśród oprogramowania dominują pakiety do analizy statystycznej, m.in.: R, SPSS, SAS. Najczęściej autorzy odwoływali się do pakietu R (ponad 20%), nieco rzadziej do SPSS oraz arkusza kalkulacyjnego Microsoft Excel. Popularność środowiska R będzie jeszcze większa, jeśli uwzględnione zostaną również pakiety rozwijające możliwości tego oprogramowania. Wtedy liczba odwołań do wspomnianego środowiska sięgnie 50%.

Badacze odwoływali się w swoich publikacjach do bardzo różnorodnych kategorii oprogramowania. Zidentyfikowano np.: systemy zarządzania bazami danych (MySQL, Microsoft Access), aplikacje do wizualizacji danych (WordArt, Wordle), oprogramowanie z grupy systemów informacji geograficznej (ArcGIS, QGIS) czy pakiety do wyszukiwania i pobierania metadanych bibliograficznych (rscopus, rcrossref, easyPubmed, rentrez). W badanym

zasobie nie zabrakło również aplikacji do analizy sieci społecznych, niewymienionych w grupie oprogramowania będącej przedmiotem analizy (Guess, Rstatnet, tidygraph). W analizach bibliometrycznych wykorzystywane jest również oprogramowanie do przetwarzania języka naturalnego (Mallet, Natural Language Toolkit).

Skala wykorzystania innego rodzaju oprogramowania w analizach bibliometrycznych niewątpliwie jest o wiele większa. Niniejsze badanie koncentruje się tylko na tej grupie publikacji, które wykorzystują oprogramowanie do analizy bibliometrycznej i do analizy sieci społecznych. Badacze rezygnują jednak niekiedy z tego rodzaju oprogramowania na rzecz innych aplikacji. Dowodzą tego statystyki wykazywania innych popularnych aplikacji. Pakiet SPSS został wymieniony w blisko 70 artykułach (jeśli nie uwzględnimy artykułów zidentyfikowanych na potrzeby niniejszego badania), Stata w prawie 40, a fraza Microsoft Excel pojawia się w blisko 30 publikacjach. Oprogramowanie do analizy danych jakościowych (NVIVO, MAXQDA) zostało wskazane przez autorów 10 publikacji.

6.3. Współwystępowanie aplikacji

Biorąc pod uwagę całość zidentyfikowanego oprogramowania, liczba artykułów, w których wykorzystano tylko jeden program wynosi 32.64%. W identycznym stosunku w artykułach wskazywano dwie aplikacje (Tab. 4).

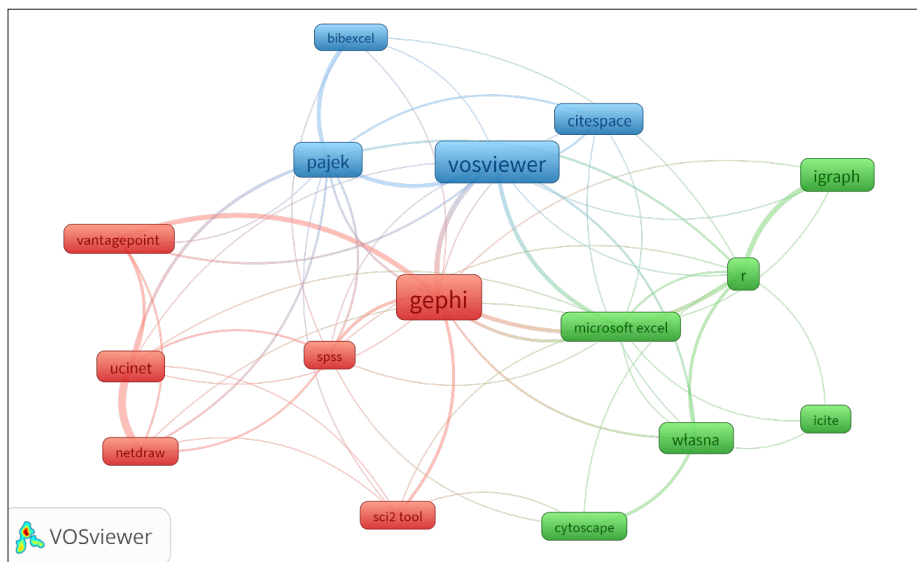
Tab. 4. Liczba aplikacji wykorzystywanych w artykułach

Liczba wykorzystywanych aplikacji	2009–2013	% (n=21)	2014–2017	% (n=59)	2018–2021	% (n=64)
1	6	28.57	19	32.20	22	34.38
2	6	28.57	19	32.20	22	34.38
3	4	19.05	12	20.34	12	18.75
4	4	19.05	7	11.86	4	6.25
5	1	4.76	0	0.00	2	3.13
6	0	0.00	1	1.69	0	0.00
7	0	0.00	1	1.69	0	0.00
8	0	0.00	0	0.00	0	0.00
9	0	0.00	0	0.00	1	1.56
10	0	0.00	0	0.00	1	1.56
Razem	21		59		64	

Program Gephi, który był najczęściej wymieniany przez autorów publikacji, jako jedyna aplikacja został wymieniony tylko w około 12% tekstów. Dla porównania – VOSviewer został wskazany jako jedyna aplikacja w 45% artykułów.

Do aplikacji, które stosunkowo często występują jako jedyne stosowane oprogramowanie należą: VOSviewer (45.16%), CiteSpace (50%), iCite (71.43%) czy bibliometrix (50%). Z kolei BibExcel, czy NetDraw są przykładami oprogramowania, które zawsze było stosowane z innymi aplikacjami.

Najczęściej cytowane Gephi było razem wymieniane z największą liczbą 41 aplikacji. Z 26 programami były wykazywane VOSviewer i Pajek. Trzy najczęściej wymieniane pary aplikacji to: UCINET i NetDraw, Gephi i VOSviewer oraz Gephi i VantagePoint. Wzajemne powiązania aplikacji, które zostały wymienione przynajmniej pięciokrotnie w analizowanych artykułach przedstawia rysunek 2.



Rys. 2. Wzajemne powiązania aplikacji wymienionych przynajmniej pięciokrotnie w analizowanych artykułach

7. Wnioski

Omówione w artykule badanie wykazało dużą różnorodność wykorzystywanego oprogramowania z wyraźną przewagą kilku aplikacji (Gephi, VOSviewer, Pajek, iGraph, UCINET, CiteSpace, VantagePoint) stosowanych stosunkowo często w całym analizowanym okresie. Rodzaj stosowanego oprogramowania pokazuje także charakter przeprowadzanych analiz. Badacze interesują się głównie relacjami zachodzącymi w nauce, współpracą badawczą, identyfikowaniem głównych nurtów badawczych. Pewne wyobrażenie o charakterze badań daje np. analiza słów występujących w tytułach publikacji. W prawie 20% artykułów pojawia się np. słowo „collaboration” czy fraza „social network”. Około 10% tekstów to analizy identyfikujące główne nurty badawcze w wybranym obszarze nauki (tytuły ze słowem „mapping”, np. „mapping the knowledge”).

Duża liczba wykorzystywanego oprogramowania jest dowodem na podejmowanie przez badaczy zróżnicowanej tematyki, do analizy której dobierane lub tworzone jest specjalne oprogramowanie. Świadczy o tym również wysoki, wynoszący blisko 70%, udział artykułów, w których wykorzystano więcej niż jedną aplikację. Badacze posilają się dodatkowym oprogramowaniem ułatwiającym zebranie i przygotowanie zebranych danych do końcowej analizy.

Niski poziom częstości wykorzystania niektórych aplikacji wynika prawdopodobnie z tego, że nie są one już rozwijane przez ich twórców. Przypadek ten dotyczy np. oprogramowania HistCite (Clarivate, 2021), CopalRed, Sitkis, czy SAINT (Hamid Jamali R., n.d.). Ponadto część programów została dopiero niedawno udostępniona (np. Ambalytics, Tethne) i jeszcze nie weszła do repertuaru narzędzi stosowanych przez badaczy. Interesujące jest też to, że mimo iż teksty reprezentują głównie obszar nauk medycznych, to badacze stosunkowo rzadko wykorzystywali oprogramowanie specjalnie zaprojektowane pod kątem medycznych analiz bibliometrycznych. Oprogramowanie iCite oraz PubReminer zostało wskazane tylko w niecałych 6% analizowanych artykułów.

Bibliografia

- Archambault, É., Campbell, D., Gingras, Y., Larivière, V. (2009). Comparing Bibliometric Statistics Obtained From the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, 60(7), 1320–1326. <https://doi.org/10.1002/asi.21062>
- Aria, M., Cuccurullo, C. (2017). bibliometrix: An R-Tool for Comprehensive Science Mapping Analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Bales, M. E., Wright, D. N., Oxley, P. R., Wheeler, T. R. (2020). *Bibliometric Visualization and Analysis Software: State of the Art, Workflows, and Best Practices* [online] [19.12.2021], <https://ecommons.cornell.edu/handle/1813/69597>
- Bouchard, L., Albertini, M., Batista, R., de Montigny, J. (2015). Research on Health Inequalities: A Bibliometric Analysis (1966–2014). *Social Science & Medicine*, 141, 100–108. <https://doi.org/10.1016/j.socscimed.2015.07.022>
- Chamberlain, S., Boettiger, C., Ram, K., rOpenSci (2021). *rplos: Interface to the Search API for "PLoS" Journals (Version 1.0.0)* [online]. Cran R-Project, [19.12.2021], <https://CRAN.R-project.org/package=rplos>
- Clarivate (2021). *HistCite: No longer in active development or officially supported* [online]. Clarivate [19.12.2021], https://support.clarivate.com/ScientificandAcademicResearch/s/article/HistCite-No-longer-in-active-development-or-officially-supported?language=en_US
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., Herrera, F. (2011). Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Cooper, I. D. (2015). Bibliometrics Basics. *Journal of the Medical Library Association : JMLA*, 103(4), 217–218. <https://doi.org/10.3163/1536-5050.103.4.013>
- CRAN (n.d.). *The Comprehensive R Archive Network* [online], [19.12.2021], <https://cran.r-project.org/>
- Cronin, B. (2015). *Getting Started in Social Network Analysis with NETDRAW* [online] [19.12.2021], <https://core.ac.uk/download/pdf/74244341.pdf>
- Decan, A., Mens, T., Claes, M., Grosjean, P. (2015). *On the Development and Distribution of R Packages* [online]. *Proceedings of the 2015 European Conference on Software Architecture Workshops* [19.12.2021], <https://doi.org/10.1145/2797433.2797476>
- Dembe, A. E., Partridge, J. S., Geist, L. C. (2011). Statistical Software Applications Used in Health Services Research: Analysis of Published Studies in the U.S. *BMC Health Services Research*, 11(1), 252. <https://doi.org/10.1186/1472-6963-11-252>
- Du, C., Cohoon, J., Lopez, P., Howison, J. (2021). Softcite Dataset: A Dataset of Software Mentions in Biomedical and Economic Research Publications. *Journal of the Association for Information Science and Technology*, 72(7), 870–884. <https://doi.org/10.1002/asi.24454>
- Du, C., Howison, J., Lopez, P. (n.d.). *Softcite: Automatic Extraction of Software Mentions in Research Literature* [online] [19.12.2021], <https://scinlp.org/history/2020/pdfs/softcite-automatic-extraction-of-software-mentions-in-research-literature.pdf>

- Elie, L., Granier, C., Rigot, S. (2021). The Different Types of Renewable Energy Finance: A Bibliometric Analysis. *Energy Economics*, 93, 104997. <https://doi.org/10.1016/j.eneco.2020.104997>
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., ... Zhang, J. (2004). Bioconductor: Open Software Development for Computational Biology and Bioinformatics. *Genome Biology*, 5(10), R80. <https://doi.org/10.1186/gb-2004-5-10-r80>
- German, D. M., Adams, B., Hassan, A. E. (2013). The Evolution of the R Software Ecosystem. *17th European Conference on Software Maintenance and Reengineering*. <https://doi.org/10.1109/CSMR.2013.33>
- Hamid Jamali, R. (n.d.). *Scientometric Portal* [online] [19.12.2021], <https://sites.google.com/site/hjamali/scientometric-portal>
- Howison, J., Bullard, J. (2016b). Software in the Scientific Literature: Problems With Seeing, Finding, and Using Software Mentioned in the Biology Literature. *Journal of the Association for Information Science and Technology*, 67(9), 2137–2155. <https://doi.org/10.1002/asi.23538>
- Howison, J., Deelman, E., McLennan, M. J., Ferreira da Silva, R., Herbsleb, J. D. (2015). Understanding the Scientific Software Ecosystem and Its Impact: Current and Future Measures. *Research Evaluation*, 24(4), 454–470. <https://doi.org/10.1093/reseval/rvv014>
- Hutchins, B. I., Baker, K. L., Davis, M. T., Diwersy, M. A., Haque, E., Harriman, R. M., ... Santangelo, G. M. (2019). The NIH Open Citation Collection: A Public Access, Broad Coverage Resource. *PLoS Biology*, 17(10), e3000385. <https://doi.org/10.1371/journal.pbio.3000385>
- Joppa, L., McInerney, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., ... Emmott, S. (2013). Troubling Trends in Scientific Software Use. *Science (New York, N.Y.)*, 340, 814–815. <https://doi.org/10.1126/science.1231535>
- Kammerer, K., Göster, M., Reichert, M., Pryss, R. (2021). Ambalytics: A Scalable and Distributed System Architecture Concept for Bibliometric Network Analyses. *Future Internet*, 13(8), 203. <https://doi.org/10.3390/fi13080203>
- Kaur, J., Hoang, D. T., Sun, X., Possamai, L., JafariAsbagh, M., Patil, S., Menczer, F. (2012a). Scholarmeter: A Social Framework for Analyzing Impact Across Disciplines. *PLoS ONE*, 7(9), e43235. <https://doi.org/10.1371/journal.pone.0043235>
- Kaur, J., Hoang, D. T., Sun, X., Possamai, L., JafariAsbagh, M., Patil, S., Menczer, F. (2012b). Scholarmeter: A Social Framework for Analyzing Impact Across Disciplines. *PLoS ONE*, 7(9), e43235. <https://doi.org/10.1371/journal.pone.0043235>
- Li, K., Yan, E. (2018). Co-mention Network of R Packages: Scientific Impact and Clustering Structure. *Journal of Informetrics*, 12(1), 87–100. <https://doi.org/10.1016/j.joi.2017.12.001>
- Li, L., Catalá-López, F., Alonso-Arroyo, A., Tian, J., Alexandre-Benavent, R., Pieper, D., Yang, K. (2016). The Global Research Collaboration of Network Meta-Analysis: A Social Network Analysis. *PLoS ONE*, 11(9), e0163239. <https://doi.org/10.1371/journal.pone.0163239>
- Majeed, S., Uzair, M., Qamar, U., Farooq, A. (2020). Social Network Analysis Visualization Tools: A Comparative Review. *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 1–6. <https://doi.org/10.1109/INMIC50486.2020.9318162>
- Mayernik, M. S., Hart, D. L., Maull, K. E., Weber, N. M. (2017). Assessing and Tracing the Outcomes and Impact of Research Infrastructures. *Journal of the Association for Information Science and Technology*, 68(6), 1341–1359. <https://doi.org/10.1002/asi.23721>
- Meho, L. I., Yang, K. (2007). Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105–2125. <https://doi.org/10.1002/asi.20677>
- Microsoft (2021). *Microsoft Excel* [online] [19.12.2021], <https://www.microsoft.com/pl-pl/microsoft-365/excel>
- Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., Cobo, M. J. (2020). Software Tools for Conducting Bibliometric Analysis in Science: An Up-To-Date Review. *El Profesional de La Información*, 29(1). <https://doi.org/10.3145/epi.2020.ene.03>

- Moral-Munoz, J. A., López-Herrera, A. G., Herrera-Viedma, E., Cobo, M. J. (2019). Science Mapping Analysis Software Tools: A Review. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (eds.), *Springer Handbook of Science and Technology Indicators* (pp. 159–185). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_7
- Pajek (2007). *WoS2Pajek. Package for Large Network Analysis* [online] [19.12.2021], <http://vlado.fmf.uni-lj.si/pub/networks/pajek/WoS2Pajek/default.htm>
- Pan, X., Yan, E., Cui, M., Hua, W. (2018). Examining the Usage, Citation, and Diffusion Patterns of Bibliometric Mapping Software: A Comparative Study of Three Tools. *Journal of Informetrics*, 12(2), 481–493. <https://doi.org/10.1016/j.joi.2018.03.005>
- Pan, X., Yan, E., Cui, M., Hua, W. (2019). How Important Is Software to Library and Information Science Research? A Content Analysis of Full-Text Publications. *Journal of Informetrics*, 13(1), 397–406. <https://doi.org/10.1016/j.joi.2019.02.002>
- Pan, X., Yan, E., Hua, W. (2016). Disciplinary Differences of Software Use and Impact in Scientific Literature. *Scientometrics*, 109(3), 1593–1610. <https://doi.org/10.1007/s11192-016-2138-4>
- Pan, X., Yan, E., Wang, Q., Hua, W. (2015). Assessing the Impact of Software on Science: A Bootstrapped Learning of Software Entities in Full-Text Papers. *Journal of Informetrics*, 9(4), 860–871. <https://doi.org/10.1016/j.joi.2015.07.012>
- PLOS (n.d.). *Text & Data Mining* [online] [19.12.2021], <https://plos.org/text-and-data-mining/>
- PLOS ONE (n.d.). *Materials, Softwares, and Code Sharing* [online]. Plos One, [19.12.2021], <https://journals.plos.org/plosone/s/materials-software-and-code-sharing>
- Pradhan, P. (2016). Science Mapping and Visualization Tools Used in Bibliometric & Scientometric Studies: An Overview. *INFLIBNET Newsletter*, 23, 19–33.
- R Foundation (2020). *The R project for statistical computing* [online]. The R Foundation, [19.12.2021], <https://www.R-project.org/>
- Roldan-Valadez, E., Salazar-Ruiz, S. Y., Ibarra-Contreras, R., Rios, C. (2019). Current Concepts on bibliometrics: A Brief Review About Impact Factor, Eigenfactor Score, CiteScore, SCImago Journal Rank, Source-Normalised Impact per Paper, H-index, and Alternative Metrics. *Irish Journal of Medical Science*, 188(3), 939–951. <https://doi.org/10.1007/s11845-018-1936-5>
- Schindler, D., Zapilko, B., Krüger, F. (2019). *SoftwareKG* [Data set]. Zenodo. <https://doi.org/10.5281/ZENODO.3571350>
- Schindler, D., Zapilko, B., Krüger, F. (2020). Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach. *The Semantic Web*, 271–286. Springer, Cham. https://doi.org/10.1007/978-3-030-49461-2_16
- Serrat, O. (2017). Social Network Analysis. In O. Serrat (ed.), *Knowledge Solutions* (pp. 39–43). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-0983-9_9
- Slater, L. (2012). PubMed PubReMiner. *Journal of the Canadian Health Libraries Association / Journal de l'Association Des Bibliothèques de La Santé Du Canada*, 33(2), 106–107. <https://doi.org/10.5596/c2012-014>
- Smith, A. M., Katz, D. S., Niemeyer, K. E., FORCE11 Software Citation Working Group. (2016). Software Citation Principles. *PeerJ Computer Science*, 2, e86. <https://doi.org/10.7717/peerj-cs.86>
- Social network (2021, December 19). Social network analysis software [online]. *Wikipedia, The Free Encyclopedia* [19.12.2021], https://en.wikipedia.org/w/index.php?title=Social_network_analysis_software&oldid=1041385709
- Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., ... Clark, T. (2015). Achieving Human and Machine Accessibility of Cited Data in Scholarly Publications. *PeerJ Computer Science*, 1, e1. <https://doi.org/10.7717/peerj-cs.1>
- Tenner, H., Thurmayer, G. R., Thurmayer, R. (2003). Data Mining With Meva in MEDLINE. In P. Perner, R. Brause, & H.-G. Holzhütter (eds.), *Medical Data Analysis* (pp. 39–46). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-39619-2_6

- Todeschini, R., Baccini, A. (2016). *Handbook of Bibliometric Indicators: Quantitative Tools for Studying and Evaluating Research*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. <https://doi.org/10.1002/9783527681969>
- VOSviewer (1.6.17). (2021). [Windows]. https://www.vosviewer.com/downloads/VOSviewer_1.6.17_exe.zip
- van Eck, N. J., Waltman, L. (2014). Visualizing Bibliometric Networks. In Y. Ding, R. Rousseau & D. Wolfram (eds.), *Measuring Scholarly Impact* (pp. 285–320). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_13
- Xu, S., Shu, G., Qiang, S., Lei, C., Xiumei, Z. (2013). Advances in Plastic and Cosmetic Surgery at Home and Abroad – A Bibliometric Analysis. *European Review for Medical and Pharmacological Sciences* [online], 17(20), 2732–2754.

Quantitative Analysis of the Use of Software Tools in Bibliometric Studies

Abstract

Purpose/Thesis: The article examines the extent to which researchers used the software in bibliometric analyses published in the Public Library of Science (PLoS).

Approach/Methods: The articles and the metadata were downloaded from PLoS. The rplots package was used to select relevant publications. The most popular software tools employed in bibliometric analysis were identified basing research literature and relevant websites. The analysis covered 52 tools: 38 designed specifically for bibliometric analysis and 14 for social network analysis.

Results and conclusions: The use of bibliometric and social network analysis software was evidenced in 144 articles. The researchers mentioned 57.69% of analyzed applications. They used 52.63% of bibliometric software tools and 71.43% social network analysis tools at least once. Gephi and VOSviewer were mentioned most ofte. These applications were cited in more than 20% of examined articles.

Originality/Value: The results indicate the importance of specific software tools for bibliometric analysis. The article identifies the most often used programs and the patterns of usage from the last decade.

Keywords

Bibliometrics. Social network analysis. Software. Software usage.

Dr ADAM JACHIMCZYK jest adiunktem w Katedrze Bibliografii i Dokumentacji na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego. Jego zainteresowania naukowe obejmują zastosowanie technologii informatycznej w działalności informacyjnej. Opublikował m.in.: Internetowe zasoby bibliograficzne instytutów badawczych oraz jednostek Polskiej Akademii Nauk (J. Franke, J. Woźniak-Kasperek, red., Bibliografia: historia, teoria, praktyka: praca zbiorowa, 2016) Web directories: selected features and their impact on directory quality (Program-Electronic Library and Information Systems, 2016, współaut. M. Chrapek, Z. Chrapek); Otwarte dane badawcze. Casus polskich instytutów badawczych (Zagadnienia Naukoznawstwa, 2015).

Kontakt z autorem:

a.jachimczyk@uw.edu.pl

Katedra Bibliografii i Dokumentacji

Wydział, Dziennikarstwa, Informacji i Bibliologii

Uniwersytet Warszawski

ul. Nowy Świat 69

00-046 Warszawa