
Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego

Łukasz Opaliński¹

Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej

Marcin Jaromin

Zakład Informatyki Chemicznej Wydziału Chemicznego Politechniki Rzeszowskiej

Jolanta Wikiera

*Pracownia Zbiorów Specjalnych i Punkt Informacji Normalizacyjnej
Biblioteki Politechniki Rzeszowskiej*

Abstrakt

Cel/teza: Celem przedstawionych badań jest empiryczne przetestowanie hipotezy o istnieniu cechy stabilności zachowań naukowców w zakresie cytowań, rozumianej jako ich względna niezmiennosc w długich okresach czasu.

Koncepcja/metody badań: W artykule podjęto próbę uzasadnienia i empirycznego ugruntowania synchronicznej metodologii badań starzenia się publikacji naukowych, w oparciu o analizę porównawczą zachowań w zakresie cytowań wykazywanych przez naukowców pracujących w obszarze inżynierii elektronicznej. W drodze analizy statystycznej zbadano rozkład prawdopodobieństwa wieku źródeł cytowanych w czterech reprezentatywnych dla dziedziny inżynierii elektronicznej periodykach naukowych i zestawiono uzyskane wyniki z rezultatami badania, które w odniesieniu do tych samych czasopism, przedstawił Emilio Matricciani w 1991 r.

Wyniki i wnioski: Potwierdzono zgodność typów rozkładów prawdopodobieństwa, którym podlegają daty wydania cytowanych materiałów źródłowych (i tym samym charakter procesu starzenia się piśmiennictwa w wybranym obszarze nauki), co zaświadcza wstępnie o względnej stabilności zachowań w zakresie cytowań uczonych publikujących na łamach zagranicznych periodyków dziedzinowych, obejmujących swoim zakresem tematycznym inżynierię elektroniczną.

Oryginalność/wartość poznawcza: W dotychczasowej literaturze przedmiotu publikowano jedynie prace, które poruszały podstawowy problem postawiony w niniejszym artykule tylko jako zagadnienie poboczne, usytuowane na marginesie głównego toku rozważań. Oryginalność zaproponowanego ujęcia omawianej tematyki zasadza się na wyraźnym wyartykułowaniu problemu, wykorzystaniu metody modelowania statystycznego do jego rozwiązania oraz wskazaniu płynących zeń implikacji, ważnych dla poznawczego uzasadnienia synchronicznej metodologii badań starzenia się literatury naukowej.

Słowa kluczowe

Bibliometria. Czasopisma naukowe. Zachowania w zakresie cytowania. Stabilność zachowania w zakresie cytowania. Starzenie się piśmiennictwa naukowego. Metody statystyczne w informatologii.

Otrzymano: 19 sierpnia 2015. Zrecenzowano: 1 grudnia 2015. Zaakceptowano: 30 grudnia 2015.

¹ Kontakt z autorami: Łukasz Opaliński <lopa@prz.edu.pl>

1. Wprowadzenie

Problematyka starzenia się publikacji naukowych jako nośników informacji naukowej stanowi ważny i ekspansywny nurt badań, zajmujący poczesne miejsce w obrębie zagadnień będących przedmiotem współczesnej bibliometrii. Nurt ten ukonstytuował się wraz z początkiem XX w., a za pierwszą pracę w historii, w której wykorzystano analizę cytowań do pomiaru tempa starzenia się piśmiennictwa, uznaje się w literaturze przedmiotu artykuł Paula L. K. Grossa i jego małżonki E. Gross, zatytułowany *College Libraries and Chemical Education*, który ukazał się w magazynie *Science* w październiku 1927 r. (zob. np.: Alvarez et al., 2000, 812; Bouabid & Larivière, 2013, 696; Huntington et al., 2006, 1841; Kurtz & Bollen, 2010, 15; Larivière et al., 2008, 288; Motylev, 1989, 97; Nicholas et al., 2005, 1443; Skalska-Zlat, 1988, 263). Badanie to było zarazem tym, które nadało impet i pchnęło myśl wielu kolejnych badaczy w kierunku pogłębiania stanu wiedzy o zjawisku starzenia się literatury fachowej i poszukiwania praktycznych obszarów jego zastosowań. Obszarami tymi są przede wszystkim bibliotekoznawstwo (zarządzanie kolekcjami bibliotek, poszukiwanie wytycznych dla polityki gromadzenia i selekcji zbiorów tradycyjnych i elektronicznych) oraz szeroko rozumiane naukoznawstwo, tj. analiza komunikacji naukowej w znaczeniu, w jakim pisała o niej Christine Borgman, a więc badania wykorzystania i rozpowszechniania informacji poprzez formalne kanały jej dystrybucji, uzyskanie wglądu w przyrost informacji naukowej w czasie czy w potrzeby informacyjne i specyfikę użytkowania informacji przez naukowców, jak również w dynamikę rozwoju rozmaitych specjalności naukowych (Borgman, 1990, 13–14; zob. też: Opaliński, 2013, 153; Stefaniak, 1999, 49–53). Do obszarów praktycznych zastosowań badań starzenia się piśmiennictwa zaliczyć należy ponadto tzw. bibliometrię historyczną, w ramach której identyfikuje się np. prace klasyczne (ponadczasowe), stale obecne w dorobku literaturowym poszczególnych dziedzin wiedzy (szczególnie interesującym przykładem tego zastosowania jest tzw. metoda spektroskopii dat wydania źródeł bibliograficznych, ang. *reference publication year spectroscopy*, zaprojektowana w 2014 r. przez Wenera Marxa, Lutz Bornmanna, Andreasa Bartha i Loeta Leydesdorffa) (Marx et al., 2014, 752; zob. też: Marx & Bornmann, 2014; Leydesdorff et al., 2014).

W perspektywie metodologicznej, istnieją dwie główne analityczne odmiany ujęcia problemu starzenia się literatury naukowej, tzw. podejście synchroniczne, nazywane też retrospektywnym oraz podejście diachroniczne, określane również mianem prospektywnego (zob. np. Burrell, 2001, 3; Diodato, 1994, 122–123; Diodato & Smith, 1993, 102; Egghe & Rousseau, 2000, 1005; Glänzel, 2004, 511–512; Line & Sandison, 1974, 286–287; Stinson & Lancaster, 1987, 65). Synchroniczna metoda badań starzenia się piśmiennictwa polega na obserwacji rozkładu dat wydania źródeł cytowanych w pewnym ustalonym korpusie literatury i obliczaniu ich średniej arytmetycznej lub mediany, albo też wykreślaniu na diagramach zależności liczby cytowań materiałów od osiągniętego przez nie wieku. Metoda diachroniczna polega natomiast na obserwacji historii cytowań, które ustalony korpus literatury otrzymuje starzejąc się, tzn. na obserwacji zmian w cytowności tego korpusu postępujących wraz z wydłużaniem się okresu obserwacji (tzw. okna cytowań, ang. *citation window*). Pierwsza metoda stanowi zatem swego rodzaju „spojrzenie wstecz” na wiek literatury wykorzystywanej przez naukowców aktywnych w określonej dziedzinie nauki w pewnym, najczęściej stosunkowo współczesnym okresie czasu, druga zaś jest niejako „spojrzeniem w przyszłość”, oglądem biegu zdarzeń, które nastąpiły po pewnym ustalonym, leżącym w przeszłości momencie czasu.

Obie metody analiz skupiają się przy tym na zliczaniu cytowań, które otrzymują dokumenty naukowe. Cytowania są z kolei zjawiskiem o podłożu i uwarunkowaniach społecznych, tzn. pochodzą od naukowców, którzy podejmują decyzje co do powoływania się w trakcie swoich badań i przygotowywania publikacji na jedne, wybrane źródła bibliograficzne i co do odrzucenia innych, uznanych przez nich przypuszczalnie za mniej wartościowe lub mniej relewantne dla ich własnych dociekań. Zachowania naukowców w zakresie cytowań (ang. *citation behaviour*) są więc ostatecznie fundamentem i wyznacznikiem tego, ile cytowań (i czy w ogóle) uzyskują poszczególne prace (artykuły, monografie, patenty, itd.), składające się na konglomerat dziedzinowego archiwum. Inaczej mówiąc, za każdym aktem cytowania skrywają się złożone procesy myślowe i emocjonalne cytujących autorów, podlegające różnym wpływom psychologicznym, socjologicznym i poznawczym, z których nie wszystkie muszą mieć cokolwiek wspólnego z motywacjami (czynnikami) czysto naukowymi, takimi jak chęć spłaty intelektualnego długu zaciągniętego, np. u pionierów danej dziedziny nauki czy „altruistyczna” chęć poszerzenia aktualnych granic ludzkiej wiedzy². Analizie i rozpoznawaniu tych czynników poświęcono wiele uwagi w literaturze z zakresu socjologii wiedzy naukowej (zob. np.: Camacho-Miñano & Núñez-Nickel, 2009; Chubin & Moitra, 1975; Cole & Cole, 1967; Cronin, 1981; Cronin, 2000; Gilbert, 1977; Latour, 1987; Law & French, 1974; Merton, 1973; Small, 2004; White & Wang, 1997; Zhang et al., 2013), ale w ramach niniejszego opracowania skoncentrowano się tylko na jednym aspekcie zachowań w zakresie cytowań, który wydaje się być główną determinantą zasadności i wiarygodności synchronicznej metodologii badań starzenia się publikacji naukowych.

Aspektem tym jest mianowicie stabilność zachowań naukowców w zakresie cytowań, rozumiana jako (przynajmniej względna) niezmiennosc tych postaw, ich niezależność od wybranego w czasie momentu badania synchronicznego, w ramach którego dąży się do ustalenia rozkładu wieku źródeł cytowanych w określonym wycinku literatury naukowej. Już w 1974 r. Maurice Line i Alexander Sandison zauważyli bowiem, że prowadząc badanie synchroniczne, badacz musi poczynić pewne założenia, które domagają się uzasadnienia empirycznego. Musi on więc założyć, że wybrany przezeń rok wydania publikacji, których bibliografie załącznikowe analizuje się pod kątem dat wydania przywołanych w nich źródeł, musi być pod każdym względem typowy i że wybór innej daty wydania materiałów stanowiących dlań materiał empiryczny nie wpłynąłby zasadniczo na uzyskane rezultaty. Założenie to nie jest przy tym ani oczywiste samo w sobie, ani też wpisane w strukturę metodologii synchronicznej w sposób, który usprawiedliwiałby przyjmowanie go „milcząco” jako aksjomatu (zob. Line & Sandison, 1974, 287). Dokładnie ten sam punkt sporny wskazał nieco później Danny Wallace (Wallace, 1986, 138). Przykładowo, badając bieżący rocznik grupy czasopism reprezentatywnych dla wybranej dziedziny wiedzy (ang. *core journals*), wynikiem badania mógłby być wskaźnik tempa starzenia się *citing* lub *cited half-life*³ wynoszący 5.5 roku. Nie mamy jednak gwarancji, że gdyby badaniu poddano inne roczniki dokładnie tych samych czasopism (wcześniejsze albo późniejsze) otrzymany

² Znacznie szersze opisy czynników wpływających na zachowania cytujących autorów odnaleźć można w licznych publikacjach, w których rozpatrywane są zagadnienia typowe dla socjologii wiedzy naukowej. Przeglądu tych zagadnień dokonali m. in.: (Bornmann & Daniel, 2008; Egghe & Rousseau, 1990, 211–227; Liu, 1993; Nicolaisen, 2007).

³ Na temat różnicy pomiędzy wskaźnikiem *citing half-life* a *cited half-life*, które są corocznie wyliczane i ogłaszane w bazie danych *Journal Citation Reports*, zob. np.: (Opaliński, 2013, 155).

rezultat byłby bliski temu wyznaczonemu w uprzednim badaniu. Mogłoby wprawdzie wydawać się, że wystarczy w tym celu sprawdzić wartości wskaźników *half-life* odpowiednich czasopism w bazie Thomson Reuters – Journal Citation Reports. Podejście to miałyby jednak przynajmniej trzy poważne mankamenty. Po pierwsze, sam wskaźnik *cited* czy *citing half-life* zapewnia jedynie „statyczny”, ograniczony opis przebiegu procesu starzenia się piśmiennictwa naukowego, który jest nieporównywalnie uboższy informacyjnie niż pełny opis długofalowego oddziaływania i charakteru recepcji treści publikacji o różnym wieku w dziedzinowym środowisku naukowym, który można uzyskać modelując starzenie się piśmiennictwa ilościowymi metodami probabilistycznymi. Po drugie, należy pamiętać, że w bazach Thomson Reuters indeksowane są jedynie wybrane periodyki naukowe, co pozostawia poza zasięgiem naszej percepcji pomijane w nich kanały komunikacji naukowej (takie jak czasopisma spoza Master Journal List, albo większość monografii naukowych i materiałów konferencyjnych). Po trzecie, zakres czasowy dostępnych w bazie InCites Journal Citation Reports wskaźników *half-life* jest „oddolnie” ograniczony i sięga jedynie do 2003 r. (*InCites...*, 2014). Uniemożliwia to zatem porównanie wartości wskaźników w dłuższych niż 12-letnie okresach (odstępach) czasu.

Podsumowując, jeżeli zachowania naukowców w zakresie cytowań literatury źródłowej są niestabilne, nie byłoby podstaw aby w obydwu (rozdzielonych od siebie pewnym okresem czasu) badaniach oczekiwać zbliżonych wyników, co dyskredytowałoby synchroniczną metodologię badań i uzależniało jej wyniki od tak przypadkowej zmiennej, jaką jest wybór czasu prowadzenia badania. Jeżeli natomiast zachowania te (przynajmniej w obrębie wybranej dziedziny nauki) cechowałyby się stabilnością, rezultatom uzyskiwanym dzięki zastosowaniu metody synchronicznej można byłoby przypisać znacznie większą wiarygodność i miarodajność, które to właściwości wpływałyby np. na możliwość zasadnej ekstrapolacji efektów badania synchronicznego poza przeanalizowane roczniki. Tym samym, wykazanie uniwersalności (stabilności) zachowań uczonych byłoby istotnym krokiem w kierunku epistemologicznego uzasadnienia wniosków wypływających z badań starzenia się piśmiennictwa naukowego w bibliotekoznawczych, informatologicznych lub naukoznawczych obszarach praktycznych zastosowań wiedzy o fenomenie spadku stopnia wykorzystania źródeł literaturowych wraz z wydłużaniem się ich wieku.

2. Cel badań

Celem naszych badań jest przetestowanie hipotezy o istnieniu cechy stabilności (trwałości) zachowań naukowców w zakresie cytowań w obszarze inżynierii elektronicznej na drodze analizy porównawczej. Podstawą doświadczalną i punktem odniesienia prezentowanych badań autorów są wyniki uzyskane przez Emilio Matriccianiego z 1988 r., które opublikowano w 1991 r. (Matricciani, 1991). Autor ten zdołał ustalić ogólny trend (wzorzec) cytowania źródeł bibliograficznych o różnych datach wydania, charakterystyczny dla autorów publikujących w czterech wiodących periodykach dziedzinowych wydawanych przez lidera światowego rynku wydawniczego w zakresie nauk elektronicznych – Institute of Electrical and Electronics Engineers (IEEE) założonego w 1884 r. Do współzałożycieli tego instytutu należą m.in. Thomas Edison i Alexander Graham Bell (*IEEE...*, 2015). Wzorzec wykryty przez Matriccianiego przybrał formę tzw. log-normalnej (logarytmicznie

normalnej) funkcji gęstości prawdopodobieństwa zapewniającej precyzyjny, ilościowy opis częstości cytowań prac o określonym wieku nie tylko w przebadanej przez autora próbie czterech reprezentatywnych czasopism, ale i w całej dziedzinowej populacji dokumentów (artykułów) naukowych (Matricciani, 1991, 10). Częstość ta (czyli stosunek liczby cytowań uzyskanych w 1988 r. przez dokumenty o wieku np. trzech, czterech, pięciu czy innej liczbie lat, do całkowitej liczby cytowań, które otrzymały wszystkie cytowane pozycje literaturowe w tym roku) może być interpretowana właśnie w kategoriach prawdopodobieństwa tego, że opis wybrany losowo z którejkolwiek bibliografii załącznikowej artykułu, który ukazał się w pewnym dziedzinowym czasopiśmie, będzie odsyłał czytelnika do publikacji wydanej przed np. trzema, czterema, lub pięcioma laty.

Materiał empiryczny, który posłużył autorom niniejszego artykułu za podstawę przeprowadzonych analiz oraz wyciągnięcie na ich podstawie wniosków został zdeponowany w serwisie Figshare⁴, który jest przeznaczony do udostępniania surowych danych badawczych. Bezpośrednie odnośniki do wszystkich trzech aneksów w formie elektronicznej znajdują się w sekcji *Bibliografia* niniejszego artykułu (zob. *Aneks 1; Aneks 2; Aneks 3*).

3. Metoda badań

Zastosowana przez autorów metoda badań jest z założenia dokładnym odwzorowaniem oryginalnej metody Matriccianiego, zastosowanej przezeń do analizy artykułów opublikowanych w 1988 r. w czterech czasopismach wydawanych przez IEEE: *Proceedings of the IEEE*, *IEEE Transactions on Communications*, *IEEE Transactions on Computers* oraz *Transactions on Lightwave Technology* (Matricciani, 1991, 7). W przypadku ostatniego tytułu można odnieść wrażenie, że jego nazwa została niewłaściwie przytoczona przez Matriccianiego, ponieważ IEEE wydaje od 1983 r. czasopismo zatytułowane *Journal of Lightwave Technology*, a tytułu rozpoczynającego się od wyrażenia *Transactions...* autorom niniejszego opracowania nie udało się odnaleźć ani w głównych, ogólnościatowych bazach danych bibliograficznych i indeksach cytowań (Web of Science, Scopus, MathSciNet Mathematical Reviews, IEEE Xplore), ani też w wyszukiwarce naukowej Google Scholar. W badaniach własnych autorów przyjęto wobec tego, że poza trzema tytułami, które wciąż ukazują się w niezmienionej formie, uzasadnione jest włączenie w obszar badania czasopisma *Journal of Lightwave Technology* jako czwartego tytułu⁵. Matricciani uwzględnił w swojej analizie jedynie artykuły badawcze (ang. *papers*), które ukazały się w czasopismach lub materiałach konferencyjnych, jako źródła informacji o datach wydania poszczególnych pozycji bibliograficznych, wyłączając z zakresu badań listy do redakcji (ang. *letters, correspondence*) i „krótkie raporty” (ang. *short papers*) (Matricciani, 1991, 7). W odniesieniu do pozycji cytowanych, Matricciani (oraz za nim autorzy niniejszej pracy) uwzględnił następujące kategorie publikacji: artykuły z czasopism, materiały konferencyjne (ang. *any written source*

⁴ Zob. <http://figshare.com/>

⁵ Dodatkowym argumentem przemawiającym za tym przypuszczeniem jest liczba artykułów badawczych w roczniku 1988 tego czasopisma. Podczas gdy Matricciani podał, że liczba publikacji typu *papers* w tym woluminie wynosiła 231 (Matricciani, 1991, 9), według danych z Web of Science, po wykluczeniu ze zbioru artykułów tych, które przyporządkowano do typów *letters, editorials* i *corrections*, pozostało ich 233 (WoS..., 2015).

published in scientific or technical journals or in conference proceedings), dysertacje (ang. *dissertation reports*) oraz książki. Ponadto, inaczej niż w badaniu Matriccianiego, zakresem analizy objęto również cytowane dokumenty elektroniczne. W niniejszym badaniu nie uwzględniono cytowań raportów technicznych, prywatnej korespondencji, patentów oraz dokumentów, w przypadku których oznaczenie odpowiedzialności wskazywało na ciała zbiorowe (Matricciani, 1991, 7). W badaniu dokonano analizy cytowań publikacji z czterech wspomnianych tytułów czasopism wydanych w 2014 r. Źródłem szczegółowych informacji o cytowaniach występujących w tym korpusie publikacji była internetowa biblioteka cyfrowa wydawcy – IEEE Xplore Digital Library.

Po dokonaniu selekcji materiału badawczego należało skupić uwagę na wielkości próby, która byłaby reprezentatywna dla dziedzinowego stylu cytowań („kultury cytowań”). Matricciani przebadał wszystkie cytowania w roczniku 1988 w czterech wybranych przez siebie tytułach czasopism, a ich suma wyniosła 14 075 pozycji (zob. Matricciani, 1991, 9). Ponieważ jednak średnia objętość bibliografii załącznikowej artykułów, a zatem i średnia liczba cytowanych przez autorów pozycji bibliograficznych, wzrasta wraz z upływem czasu (np. Matthew Wallace, Vincent Larivière i Yves Gingras stwierdzili, że w latach 1980–2004 średnia objętość bibliografii załącznikowych w ogólnościowym systemie nauki podwoiła się – zob. Wallace et al., 2009, 300), dokładna rekonstrukcja tego aspektu metody Matriccianiego wymagałoby przeanalizowania około 40 tys. odesłań bibliograficznych⁶. Z drugiej strony, podręczniki statystyki nie podają minimalnej wielkości próby, wystarczającej do przeprowadzenia tzw. testu zgodności/dokładności dopasowania rozkładów chi-kwadrat (ang. *chi-square goodness-of-fit test*), który – wzorem metodologii Matriccianiego – zostanie wykorzystany i krótko wyjaśniony w dalszej części. Autorzy podręczników statystyki posługują się natomiast tym testem nawet wtedy, kiedy próba, którą dysponują jest mniej liczna (obejmuje np. mniej lub niewiele ponad 100 wyników doświadczenia losowego) (zob. np.: Sheskin, 2007, 259; Snarska, 2011, 114–117; Taylor, 2012, 307–308; Vaughan, 2008, 76). Problem ten poruszył m.in. John Taylor twierdząc, że testu tego nie da się sensownie zastosować, kiedy liczba wszystkich posiadanych przez badacza pomiarów (wyników doświadczenia) jest mniejsza niż 20 (Taylor, 2012, 314). Z drugiej strony, John McDonald zarekomendował dla testu próbę sięgającą 1000 obserwacji (McDonald, 2014), co obrazuje też brak jednej, ścisłej reguły rządzącej rozmiarem badanej (testowanej) zbiorowości. Wobec tego wydaje się, że zebranie danych o co piętnastym opisie z bibliografii załącznikowych znajdujących się w roczniku 2014 wybranych czasopism, co da liczbę wyników doświadczenia losowego przekraczającą 2500, powinno zapewnić odpowiednią i „bezpieczną” podstawę empiryczną do przeprowadzenia testu.

W obrębie zbadanej próby odnotowano miesiąc i rok wydania co piętnastej publikacji odnalezionej w bibliografiach załącznikowych artykułów należących do wyselekcjonowanego zbioru. Wynik tego zabiegu przedstawia tabela zamieszczona w aneksie (zob. *Aneks I*). W przypadku analizy cytowań materiałów konferencyjnych najczęściej odnotowywano miesiąc, w którym odbyła się konferencja a nie miesiąc, w którym ukazała się publikacja.

⁶ W czterech wskazanych tytułach czasopism, które poddano badaniu, w 2014 r. opublikowano łącznie 1318 tekstów. Przyjmując, że w każdym z nich zamieszczono średnio około 30 cytowań (na co wskazuje przejrzanie kilku losowo wybranych zeszytów czasopism), daje to sumaryczną liczbę 39540 pozycji bibliograficznych.

Informacje te były często dostępne w treści odnośnych odesłań bibliograficznych. W przypadkach, w których nie podano dokładnego miesiąca publikacji cytowanego materiału źródłowego, autorzy podjęli próbę ustalenia tej daty w oparciu o dostępne źródła internetowe. Kiedy próba ta okazywała się nieskuteczna, jako miesiąc wydania przyjmowano czerwiec (tj. przyjmowano, że publikacja ukazała się w połowie roku). Najczęściej sytuacja ta miała miejsce w przypadku wydawnictw zwartych oraz materiałów wydanych w latach 60. i 70. XX w. Wiek źródła w momencie cytowania jest rozumiany intuicyjnie („zdroworozsądkowo”) i oznacza wyrażoną w miesiącach różnicę pomiędzy miesiącem wydania a miesiącem zacytowania danej pozycji bibliograficznej. Do jego obliczenia wykorzystano funkcję programu Microsoft Excel 2007, występującą w nim (oraz w innych wersjach tego programu) pod nazwą „DATA.RÓŻNICA(...)” (zob. np.: Walkenbach, 2011, 247). Podaje ona w wyniku liczbę pełnych miesięcy, które upłynęły między dwiema dowolnymi datami. Na przykład wynik formuły dla dat: „lipiec 2013” (w formacie MS Excel: „lip-13”) i „styczeń 2014” (w formacie MS Excel: „sty-14”) to sześć miesięcy. Ustalono w ten sposób wartości zamieniono na odpowiadające im liczby lat, dzieląc w każdym przypadku uzyskaną wartość przez 12. Nie zastosowano formuły MS Excel wyznaczającej od razu liczbę pełnych lat pomiędzy dwiema datami ze względu na to, że wartość ta zostałaby wtedy zaokrąglona do najbliższej wynikowi liczby całkowitej.

Kolejnym etapem badawczym było obliczenie niezbędnych parametrów rozkładu – średniej arytmetycznej („ μ ”) i odchylenia standardowego („ σ ”) dla **logarytmów** zaobserwowanych wartości wieku źródeł cytowanych w artykułach z czasopism (zob. Matricciani, 1991, 12). Przykładowo, logarytmem naturalnym wieku źródła równego trzy lata: „ $\ln_e 3$ ” jest liczba 1.099, ponieważ $e^{1.099} \approx 3$. Po zamianie wszystkich uzyskanych w badaniu danych dotyczących wieku cytowanych źródeł na logarytmy tej wartości (co można osiągnąć np. dzięki funkcji programu MS Excel występującej pod nazwą „LN(...)”) i wyznaczeniu powyższych parametrów, można było przejść do ostatniego etapu badania, czyli przetestowania zgodności rozkładu zaobserwowanego doświadczalnie i teoretycznego rozkładu log-normalnego, poprzez porównanie ich za pomocą testu chi-kwadrat (Matricciani, 1991, 9).

Test chi-kwadrat (χ^2) jest jednym z najpowszechniej stosowanych i najbardziej znanych testów zgodności dopasowań rozkładów teoretycznych do danych empirycznych. Jego idea zasadza się na podzieleniu zaobserwowanych danych na określoną liczbę przedziałów i porównaniu częstości danych w tych przedziałach („gęstości zaludnienia” przedziałów, licznosci wyników, które trafiły do każdego z nich) z częstościami (licznosciami) przewidywanymi, tzn. częstościami, które powinniśmy zaobserwować, gdyby istotnie rozkład stwierdzony empirycznie był taki sam, jak rozkład teoretyczny, o którym zakładamy, że leży u podstaw obserwowanej zmienności danych i którego „ukrytą obecność” w naszych danych „podejrzewamy” i sprawdzamy. Wynik testu pozwala na zaakceptowanie lub odrzucenie tzw. hipotezy zerowej (ang. *null hypothesis*), która w ogólności stwierdza brak różnicy pomiędzy badanymi zbiorowościami lub ich wybranymi parametrami (zob. np.: Sheskin, 2007, 57, 259–261; Snarska, 2011, 114–117; Taylor, 2012, 307–322; Vaughan, 2008, 76–84). W rozpatrywanym aktualnie przypadku hipoteza zerowa będzie zatem mówić, że istnieje zgodność (identyczność) rozkładu zaobserwowanego w badaniu z teoretycznym rozkładem log-normalnym. Wynik testu zawsze przybiera postać wartości liczbowej, tzw. statystyki testowej (ang. *test statistic*) którą ocenia się poprzez porównanie jej z wartościami określonymi jako krytyczne (ang. *critical values*), zawartymi w specjalnych tabelach.

Pozwalają one na podjęcie decyzji co do możliwości zaakceptowania lub odrzucenia hipotezy zerowej.

Architektura testu chi-kwadrat opiera się na oszacowaniu statystycznej istotności różnicy pomiędzy wynikami zaobserwowanymi a przewidywanymi, czyli takimi, jakich spodziewalibyśmy się w sytuacji, w której hipoteza zerowa jest prawdziwa. Im większa różnica pojawi się między obiema grupami wyników, tym mniej prawdopodobne jest, że przyczyną tej różnicy jest przypadek (Sheskin, 2007, 59–60). Kiedy badacz mówi, że różnica jest statystycznie istotna (ang. *statistically significant*), daje tym samym do zrozumienia, że jest wysoce nieprawdopodobne, iż różnica równa (lub większa) od zaobserwowanej mogła być wynikiem przypadkowych, statystycznych fluktuacji. W statystyce obowiązuje konwencja, według której uznanie różnicy za statystycznie istotną wymaga, aby ryzyko jej przypadkowości nie przekraczało 5% (Sheskin, 2007, 60). Jeżeli więc wynik testu jest statystycznie istotny (czyli istnieje mniej niż 5% szans na to, że zaobserwowana różnica jest spowodowana przypadkiem, co w formalnej notacji zapisuje się jako $p < 0.05$), należy hipotezę zerową odrzucić. Jeżeli natomiast szanse na to, że odnotowana w badaniu różnica (np. między rozkładem log-normalnym a rozkładem stwierdzonym doświadczalnie) jest dziełem przypadku są większe niż 5% (zapis: $p > 0.05$) mówimy, że wynik testu nie jest istotny na poziomie 5% (Sheskin, 2007, 60). Różnica jest wtedy – wedle wszelkiego prawdopodobieństwa – spowodowana przypadkowością rozkładu danych w próbie, co pociąga za sobą konieczność zaakceptowania hipotezy zerowej. Podsumowując, wynik testu jest nieistotny (ang. *insignificant*) kiedy różnica między danymi jest przypadkowa. Hipoteza zerowa zostaje wobec tego zaakceptowana (najczęściej na poziomie 5% istotności, chociaż możliwe jest też przyjęcie bardziej restrykcyjnego kryterium 1% istotności). Wynik testu jest natomiast istotny, kiedy różnica jest nieprzypadkowa. Hipoteza zerowa zostaje wtedy odrzucona na poziomie 5% istotności.

O tym, która sytuacja zachodzi w konkretnym badaniu, gdy dysponuje się rzeczywistymi danymi pomiarowymi, można przekonać się sięgając do tablic rozkładu chi-kwadrat (zob. np. Sheskin, 2007, 1661). Jeżeli wynik testu chi-kwadrat (statystyka testowa) będzie liczbą większą lub równą wartości krytycznej, hipoteza zerowa powinna zostać odrzucona. Można wtedy powiedzieć, że z prawdopodobieństwem 95% rozkład doświadczalny (lub jego pewna cecha, parametr) nie jest zakładanym rozkładem teoretycznym, ponieważ obserwowane odchylenia od „idealnych” częstości wyników, przewidywanych przez rozkład teoretyczny, są zbyt duże, by można było przypisać je przypadkowi. Jeżeli zaś wynik testu jest liczbą mniejszą niż wartość krytyczna, hipoteza zerowa (o istnieniu zgodności rozkładów) powinna zostać zaakceptowana (Sheskin, 2007, 261). W przeprowadzonych badaniach przyjęto, że hipoteza zerowa wyraża istnienie zgodności rozkładu dat wydania źródeł cytowanych w publikacjach z czterech wybranych tytułów czasopism z rozkładem log-normalnym o wyznaczonych (na podstawie danych empirycznych) parametrach. Jeżeli wynik testu będzie przemawiał za możliwością zaakceptowania tej hipotezy, jego interpretacja polegać będzie na sformułowaniu wniosku o istnieniu cechy stabilności zachowań naukowców w zakresie cytowań w obrębie pola tematycznego inżynierii elektronicznej. W przeciwnym razie interpretacja wyniku testu będzie z konieczności zmierzać w kierunku odrzucenia wniosku o istnieniu tej cechy wśród autorów pracujących w zbadanym obszarze nauki.

4. Wyniki badań

Po przeanalizowaniu dat wydania 2939 cytowanych publikacji uzyskano dane, które charakteryzują specyfikę cytowanych publikacji źródłowych (a zatem także – pośrednio – zachowania w zakresie cytowań wybranej grupy uczonych), oraz których część posłużyła do przeprowadzenia testu chi-kwadrat (zob. Tabela 1).

Tabela 1. Parametry i ich wartości opisujące próbę badawczą

Parametr	Wartość parametru
Liczebność próby	2939
Wartość minimalna (w latach)	0.25
Wartość maksymalna (w latach)	64.17
Średnia arytmetyczna (w latach)	9.054
Mediana (w latach)	5.92
Dominanta (lub „moda”: wartość występująca najczęściej, podana w latach)	1.92
Odchylenie standardowe (w latach)	9.662
Średnia logarytmów dat wydania źródeł cytowanych (parametr μ rozkładu)	1.735
Odchylenie standardowe wartości logarytmów wieku źródeł cytowanych (parametr σ rozkładu)	1.009
Liczba przedziałów, na które podzielono cały zbiór danych	21
Szerokość przedziału (w latach)	3.2

W następnej kolejności sprawdzono częstość, z jaką wyniki rozłożyły się w 21 przedziałach o równej szerokości. Krańce przedziałów oznaczają granice wiekowe, jakich należące do przedziału daty wydania źródeł cytowanych nie mogą przekraczać. Wizualizację rozkładu częstości w przedziałach przedstawiono na histogramie, który skonstruowano z wykorzystaniem dodatku „Analiza danych” programu MS Excel 2007 (Rys. 1).

Słupek histogramu podpisany jako „0.25” oznacza przedział, którego dolną granicą jest wiek równy 0 lat, a granicą górną wiek równy 0.25 roku. Słupek podpisany jako „3.45” oznacza przedział, którego dolną granicą jest wiek równy 0.25 roku, a granicą górną 3.45 roku, itd. Ponieważ kształt histogramu przypomina wyraźnie charakter rozkładu danych obserwowany w przypadkach, w których podlegają one rozkładowi log-normalnemu, przeprowadzono dla nich test zgodności rozkładów chi-kwadrat, ściśle według procedury streszczonej przez Agnieszkę Snarską (Snarska, 2011, 114–117). Jak już wspomniano, test ten opiera się na porównaniu empirycznych i teoretycznych (oczekiwanych) częstości, z jaką występują wyniki doświadczenia w z góry określonych przedziałach. Liczba przedziałów, podobnie jak niezbędna wielkość próby nie jest w podręcznikach statystyki matematycznej dokładnie określona, z tym jedynie wyjątkiem, że nie może ich być mniej niż cztery. Jeżeli przedziałów jest więcej, test staje się bardziej „wrażliwy” (czuły na występujące odchylenia od wartości oczekiwanych). Wybrana przez autorów liczba przedziałów równa 21 może z pewnością zostać uznana za dużą i zapewniającą wysoką precyzję analizy. Matricciani

nie podał zastosowanej przez siebie liczby przedziałów, w związku z czym w tym aspekcie badania autorzy zmuszeni byli do polegania na własnym osądzie. Uzyskane w badaniu częstości empiryczne i częstości oczekiwane przybrały formy, które przedstawiono w Tabeli 2.

Rys. 1. Histogram przedstawiający rozkład wiekowy cytowanych publikacji

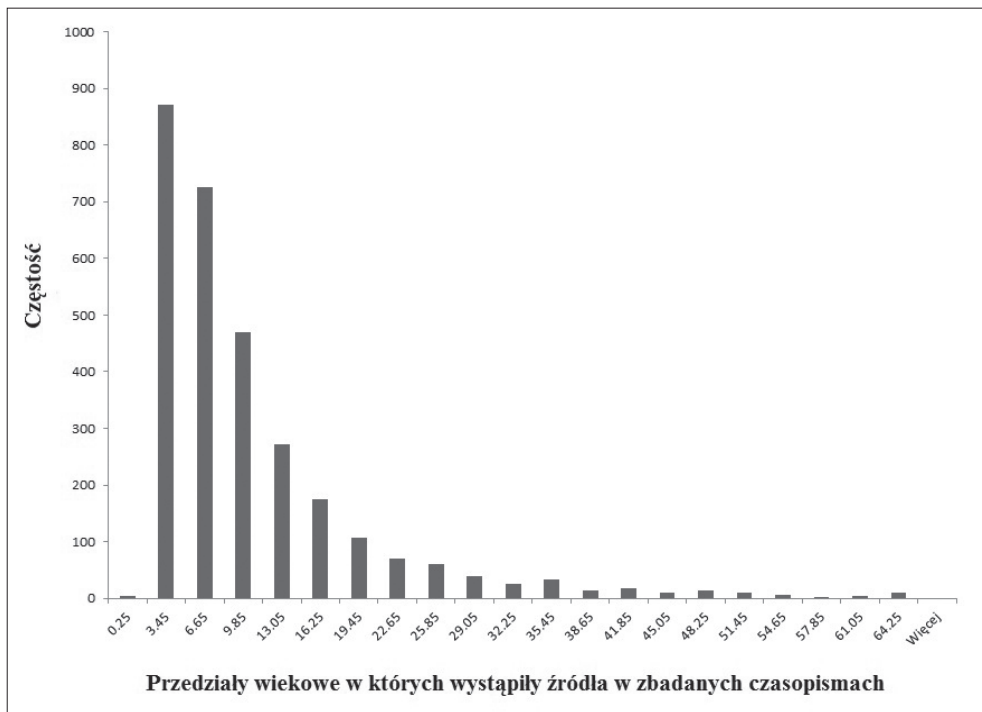


Tabela 2. Porównanie częstości zaobserwowanych w badaniu z częstościami oczekiwany dla rozkładu log-normalnego o parametrach $\mu = 1.735$ i $\sigma = 1.009$

Przedział wiekowy	Częstość zaobserwowana	Zaokrąglona częstość oczekiwana
0 – 0.25	5	2.90
0.25 – 3.45	872	911.97
3.45 – 6.65	726	739.33
6.65 – 9.85	469	426.59
9.85 – 13.05	271	257.75
13.05 – 16.25	175	164.55
16.25 – 19.45	106	110.00
19.45 – 22.65	71	76.34
22.65 – 25.85	60	54.63

25.85 – 29.05	39	40.10
29.05 – 32.25	25	30.07
32.25 – 35.45	33	22.97
35.45 – 38.65	14	17.83
38.65 – 41.85	17	14.04
41.85 – 45.05	10	11.19
45.05 – 48.25	14	9.02
48.25 – 51.45	9	7.34
51.45 – 54.65	6	6.02
54.65 – 57.85	3	4.98
57.85 – 61.05	5	4.16
61.05 – 64.25	9	3.49
> 64.25	0	23.69

Pomimo niewielkich rozbieżności pomiędzy danymi empirycznymi a oczekiwanymi, wyniki testu chi-kwadrat nakazały jednak odrzucić hipotezę o identyczności rozkładu zaobserwowanego w badaniu z rozkładem log-normalnym. Jego wynik (statystyka testowa) wyniósł w zaokrągleniu 53.42 (zob. *Aneks 2*), co według tablic prawdopodobieństw dla testu jest wartością większą od wartości krytycznej⁷, która wynosi 28.87 (zob. Sheskin, 2007, 1661). Oznacza to, iż różnica między danymi jest zbyt duża, aby można było uznać ją za przypadek przy założonym rozkładzie. Zachodzi zatem konieczność odrzucenia przewidywanego rozkładu na poziomie 5% istotności.

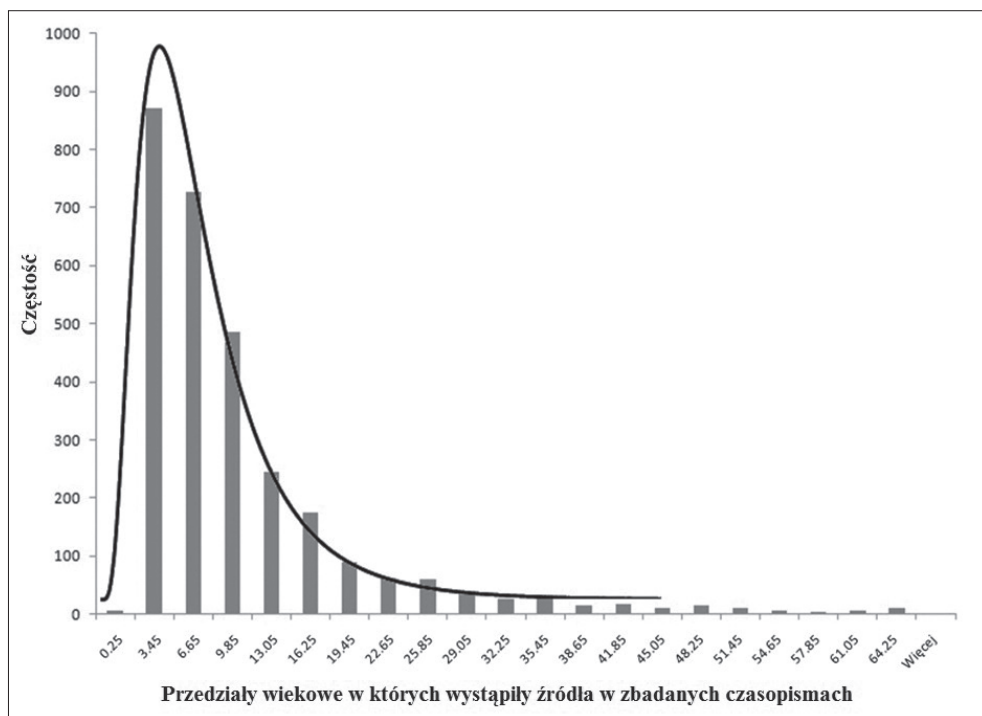
Nałożenie wykresu funkcji log-normalnej na histogram częstości występowania źródeł o określonym wieku w zbędnych bibliografiach załącznikowych przedstawia rysunek 2.

Z formalnego punktu widzenia wynik testu chi-kwadrat nie pozostawia wątpliwości co do konieczności odrzucenia hipotezy wyjściowej. Kiedy jednak zwróci się uwagę na to, które przedziały wieku źródeł cytowanych wykazały największą rozbieżność pomiędzy doświadczeniem a przewidywaniami, nietrudno zauważyć że są nimi przedziały, w których zidentyfikowano cytowanie publikacji najbardziej „wiekowych”, tj. przedział (61.05 – 64.25) i przedział (> 64.25). To z nimi bowiem wiążą się najwyższe wartości statystyki testowej wskazujące na „newralgiczne” punkty odpowiedzialne w największym stopniu za przekroczenie wartości krytycznej przez statystykę testową (zob. *Aneks 2*). Autorzy powzięli wobec tego podejrzenie, że wykluczenie z analizy prac starszych niż sześćdziesięcioletnie (cytowanych w znikomej ilości) wpłynęłoby na wynik testu i pozwoliłoby na potwierdzenie hipotezy o log-normalnym rozkładzie wieku cytowanego piśmiennictwa w rozpatrywanym obszarze komunikacji naukowej. W szczególności wysoka, oczekiwana liczba wyników w ostatnim z przedziałów (> 64.25), której nie udało się zaobserwować,

⁷ Przy 18 stopniach swobody. Stopnie swobody (ang. *degrees of freedom*) to liczba przedziałów, na jakie podzielono dane, pomniejszona o liczbę tzw. węzłów, „więzi”, jakie nakłada się na dane empiryczne „odbierając im swobodę”. Pierwszym węzłem jest samo określenie liczby przedziałów, w których grupuje się dane, kolejne dwa węzły to parametry „ μ ” i „ σ ”, które wyznacza się na podstawie zbioru badanych wartości liczbowych (Taylor, 2012, 316–319).

wyduje się być swego rodzaju matematyczną „aberracją”. Oznacza to, że wykres funkcji, do której dopasowanie testowano, biegnie w nieskończoność. Innymi słowy, funkcja przyjmuje wartości dodatnie nawet dla punktów leżących na osi poziomej nieskończenie daleko od początku układu współrzędnych (czemu odpowiadają lata wydania „nieskończenie” odległe od momentu umownego „teraz”). Po zsumowaniu wszystkich wartości dodatnich, przy założonym rozkładzie uzyska się wartość większą od zera (a konkretnie liczbę równą w przybliżeniu 23.7). Dlatego celowe w tej sytuacji wydaje się ponowne przeprowadzenie testu wykluczając tym razem publikacje wydane przed 60 laty i wcześniejsze, kiedy za punkt odniesienia (moment „0”) obierze się moment badania synchronicznego (a więc rok 2014). Istotnie, przy tych założeniach, wynikiem testu chi-kwadrat jest statystyka chi równa około 25.95 (zob. *Aneks 3*). Jest ona mniejsza niż wartość krytyczna⁸ (28.87), w związku z czym uzasadnione jest uznanie zadowalającej zgodności rozkładów empirycznego i teoretycznego. Konkludując, w kontekście cytowań publikacji nie starszych niż 60 lat, zachowania naukowców – specjalistów w dziedzinie inżynierii elektronicznej – w zakresie cytowań, okazały się względnie stabilne w latach 1988–2014.

Rys. 2. Histogram z zaznaczonym przebiegiem funkcji log-normalnej o ustalonych doświadczalnie parametrach. Wykres uwidacznia niewielkie niezgodności dopasowania danych empirycznych do teoretycznych



⁸ Przy tej samej liczbie 18 stopni swobody.

5. Wnioski

W wyniku przeprowadzonego badania uzyskano rezultaty, które skorelowano z wnioskami z badań Matriccianiego (1991). Zestawienie to przedstawia Tabela 3.

Tabela 3. Porównanie wyników Matriccianiego z uzyskanymi wynikami badań własnych

Parametr	Rezultat Matriccianiego*	Rezultat badania autorów sprzed wykluczenia z zakresu badania prac starszych niż sześćdziesięcioletnie	Rezultat badania autorów po wykluczeniu z zakresu badania prac starszych niż sześćdziesięcioletnie
Średnia (μ)	1.6	1.735	1.725
Odchylenie standardowe (σ)	0.866	1.009	1.000
Mediana wieku prac cytowanych	4.95	5.92	5.92
Dominanta wieku prac cytowanych	2.34	1.92	1.92

* Parametry „ μ ” i „ σ ” zostały podane przez Matriccianiego (Matricciani, 1991, 10), pozostałe wskaźniki (mediana i dominanta) obliczono na podstawie podanych przez tego autora wzorów służących do przekształceń parametrów rozkładu do wartości opisujących bezpośrednio wiek źródeł cytowanych (Matricciani, 1991, 12).

Przeprowadzona analiza wykazała, że na przyjętym poziomie agregacji danych (którego jednostką jest grupa czasopism), zachowania badaczy w zakresie cytowań z dość dużą dokładnością odpowiadają tym, które Matricciani odnotował przed 26 laty. Najwyraźniejszym odstępstwem od wzorca z 1988 r. jest przesunięcie się mediany wieku prac cytowanych (która jest ekwiwalentem indeksu *citing half-life*) o niecały rok w kierunku publikacji starszych. Obserwacja ta pozostaje w zgodzie z ogólnoswiatowym trendem wyrażającym się w fakcie, że społeczność naukowców (pracujących we wszystkich dziedzinach nauki) polega w swojej praktyce naukowej i publikacyjnej na coraz starszym korpusie prac dziedzinowych, począwszy mniej więcej od lat 80. XX w. Zjawisko to zostało zauważone m.in. przez Vincenta Larivièrego, Èrica Archambaulta i Yvesa Gingrasa w 2008 r. (Larivière et al., 2008, 295) oraz Leo Egghego (Egghe, 2010, 243–244, 247).

Drugim elementem wymagającym ustosunkowania się w kontekście zestawienia rezultatów badania własnego z wynikami Matriccianiego jest przedstawiona przez niego konceptualizacja rozkładu log-normalnego w informatologii, nawiązująca bezpośrednio do zachowań w zakresie cytowań. Zdaniem tego autora, dobre dopasowania do danych empirycznych zapewnione przez modelowanie danych o cytawalności jako podlegających rozkładowi log-normalnemu można wyjaśnić przyjmując, że prawdopodobieństwo wybrania przez autora cytującego pewnej publikacji o określonym wieku jako „przedmiotu cytowania”, może zostać zwiększone lub zmniejszone przez pewne czynniki multiplikatywne (ang. *multiplying factors*), takie jak jakość i wartość poznawcza dokumentu potencjalnie

cytowalnego, ranga czasopisma w którym dokument ten się ukazał, społeczne powiązania autora cytującego z autorem dokumentu potencjalnie cytowanego, afiliacja autora dokumentu potencjalnie cytowanego, wiek i długość aktywności zawodowej w nauce autora cytującego, itd. (Matricciani, 1991, 9). W myśl tej koncepcji dokładny rozmiar, liczba i siła wpływu wywierana na decyzję autora cytującego przez sumę takich czynników wprawdzie nie jest znana, ale jeżeli czynników tych jest dużo, są one wzajemnie niezależne i żaden nie dominuje, należy spodziewać się, że ich „wypadkowa” (ich połączone oddziaływanie na myśl autora cytującego) poskutkuje log-normalnym ukształtowaniem się („rozłożeniem się”) wieku prac cytowanych w dziedzinowym korpusie literaturowym (Matricciani, 1991, 9). Matricciani potwierdził swoją hipotezę analizując nie tylko wspomniane cztery amerykańskie czasopisma z zakresu inżynierii elektronicznej, ale także w drodze analizy 58-letniej historii włoskiego czasopisma o analogicznym zasięgu tematycznym – *Alta Frequenza*, wzbogaconej w dalszej kolejności o analizę cytowań w wydanej w 1965 r. monografii pt. *The Cambridge Economic History of Europe*, Vol. VI: *The Industrial Revolution and After*. Ostatecznym wnioskiem Matriccianiego było stwierdzenie, że rozkład log-normalny jest na tyle uniwersalny, że można założyć, iż jest on trafną reprezentacją sposobu dystrybucji wieku prac cytowanych, która obowiązuje we wszelkich dziedzinach wiedzy⁹, a dodatkowo że w naukach inżynieryjnych ten wzorzec zachowań badaczy jest stabilny w bardzo długich okresach czasu (Matricciani, 1991, 11). Matricciani napisał jednak swój tekst w czasach gdy wolny dostęp do prac naukowych umieszczonych w wersji elektronicznej w Internecie jeszcze nie istniał. Obecnie sytuacja zmieniła się radykalnie – w środowisku sieciowym z dnia na dzień przybywa wiele prac naukowych, które funkcjonują tu w tzw. wolnym dostępie (ang. *open access*). Dzięki wyszukiwarkom takim jak m.in. Google Scholar, badacze zainteresowani dowolnym tematem uzyskują bezpośredni dostęp do wielu potencjalnie przydatnych prac naukowych. Czy wobec tego podczas przygotowywania publikacji i selekcjonowania prac innych badaczy celem zacytowania odpowiednich źródeł nie będą oni faworyzowali tego, co jest łatwiej i szybciej dostępne? Coraz więcej doniesień wskazuje, że tak właśnie się dzieje, że prace obecne w Internecie w wolnym dostępie pozyskują więcej cytowań niż te, które dostępne są na innych (przede wszystkim komercyjnych) zasadach (zob. np.: Gargouri et al., 2010; Norris et al., 2008). Tak więc współcześnie pojawił się czynnik, który może w sposób istotny wpływać na prawdopodobieństwo wyboru konkretnej pracy przez autora cytującego, czynnik niejednokrotnie zdecydowanie **dominujący** nad innymi. Czy w takiej sytuacji w dalszym ciągu prawdziwa jest teza Matriccianiego, że należy spodziewać się, iż „wypadkowa” zbioru takich nie do końca rozpoznanych czynników leży u podstaw log-normalnego rozkładu wieku prac cytowanych w dziedzinowym „archiwum” literaturowym? Problem ten wymaga z pewnością dalszych i nieco inaczej skonstruowanych dociekań empirycznych, zanim zaakceptuje się tezę Matriccianiego. Wydaje się przy tym, że zarówno rezultaty Matriccianiego, jak i wyniki przedstawionego badania własnego nie zależą jednak od prawdziwości tej tezy, ponieważ nie warunkuje ona metodologii badawczej, a jedynie stanowi wyjaśnienie, dlaczego taki akurat rozkład wieku

⁹ Warto w tym miejscu nadmienić, że rozkład wieku prac cytowanych w literaturze naukowej z powodzeniem modelowali przy użyciu rozkładu log-normalnego także Quentin Burrell (Burrell, 2002, 284), Leo Egghe i Inna Kedage Ravichandra Rao (Egghe & Rao, 1992) oraz Michael Stringer, Marta Sales-Padro i Luís Nunez-Amaral (Stringer et al., 2010, 1379, 1384).

publikacji cytowanych został w badaniach tych wykryty¹⁰. Z drugiej strony to, czy pozytywny rezultat przedstawionego badania własnego autorów (tj. możliwość zaakceptowania hipotezy o log-normalnym rozkładzie wieku źródeł cytowanych w przeanalizowanych czasopismach) stanowi wystarczające potwierdzenie słuszności powyższej tezy pozostaje na obecnym etapie kwestią problematyczną, wymagającą kolejnych badań o większej reprezentatywności i generalizowalności.

Bez wątplenia najważniejszym wnioskiem płynącym z przedstawionych rozważań jest jednak samo stwierdzenie, że proces starzenia się publikacji naukowych w obszarze inżynierii elektronicznej rzeczywiście podlega obecnie (tj. w 2014 r.) tej samej prawidłowości, której istnienie wywiódł empirycznie Matricciani w 1991 r. Można więc założyć, że na podstawie typu zaobserwowanego rozkładu prawdopodobieństwa wieku publikacji cytowanych można domniemywać o faktycznym występowaniu cechy stabilności zachowań w zakresie cytowań ze znacznie większą dozą pewności, niż byłoby to możliwe po prostym porównaniu wartości indeksów *cited* czy *citing half-life* ISI, dla wybranych periodyków. Rozkład prawdopodobieństwa zapewnia bowiem znacznie większe „bogactwo informacyjne” analiz dzięki temu, że dostarcza wiedzy o całości cyklu życiowego publikacji naukowych, umożliwia prześledzenie częstotliwości cytowań prac o dowolnie wybranym wieku oraz stwarza okazję do uzyskania wieloaspektowego wglądu w preferencje publikujących autorów i naturę dokonywanych przez nich wyborów. Co więcej, analiza zorientowana probabilistycznie wydaje się stanowić szczególnie wiarygodną technikę badań doświadczalnych, podejmowanych z zamiarem metodologicznego ugruntowania (bądź też zakwestionowania) synchronicznego podejścia do zgłębiania zjawiska starzenia się publikacji naukowych.

Bibliografia

- Alvarez, P.; Escalona, I.; Pulgarin, A. (2000). What is wrong with obsolescence? *Journal of the American Society for Information Science*, 51(9), 812–815.
- Borgman, Ch. L. (1990). Editor's Introduction. In: Ch. Borgman (ed.) *Scholarly communication and bibliometrics*. Newbury Park: Sage Publications, 10–27.
- Bornmann, L.; Daniel, H-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80.
- Bouabid, H.; Larivière, V. (2013). The lengthening of papers' life expectancy: a diachronous analysis. *Scientometrics*, 97(3), 695–717.
- Burrell, Q. (2001). Stochastic modelling of the first-citation distribution. *Scientometrics*, 52, 3–12.

¹⁰ Innymi słowy, w ramach badania Matriccianiego oraz badania autorów niniejszego artykułu nie przyjmowano tezy, że jeżeli zachowania naukowców w zakresie cytowań są określane i determinowane przez zespół nie do końca rozpoznanych, wzajemnie niezależnych czynników, wśród których nie można wskazać jednego dominującego wyrażnie nad pozostałymi, to z konieczności warunki te zawsze narzucą na rozkład wieku źródeł przez tych naukowców cytowanych formę rozkładu log-normalnego. Teza Matriccianiego mówiła raczej, że jeżeli uda się zaobserwować ten właśnie (log-normalny) rozkład wieku publikacji cytowanych w jakimś zbiorze literatury, to prawdopodobnym wytłumaczeniem tego faktu może być powyższa charakterystyka zespołu czynników oddziałujących na zachowania autorów cytujących. Możliwa jest zatem sytuacja, w której empiryczny rozkład prawdopodobieństwa wieku materiałów cytowanych jest istotnie rozkładem log-normalnym, a zarazem zachowania uczonych w zakresie cytowań nie są precyzyjnie opisywane przez wspomnianą konceptualizację (jest ona nieściśła, niepełna lub wręcz nieprawdziwa).

- Burrell, Q. L. (2002). Modelling citation age data: simple graphical methods from reliability theory. *Scientometrics*, 55(2), 273–285.
- Camacho-Miñano, M.; Núñez-Nickel, M. (2009). The multilayered nature of reference selection. *Journal of the American Society for Information Science and Technology*, 60(4), 754–777.
- Chubin, D. E.; Moitra, S. D. (1975). Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423–441.
- Cole, J.; Cole, S. (1967). Scientific output and recognition: a study in the operation of the reward system in science. *American Sociological Review*, 32(3), 377–390.
- Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation*, 37(1), 16–24.
- Cronin, B. (2000). Semiotics and evaluative bibliometrics. *Journal of Documentation*, 56(4), 440–453.
- Diodato, V. (1994). *Dictionary of bibliometrics*. New York: Haworth Press.
- Diodato, V.; Smith, F. (1993). Obsolescence of music literature. *Journal of the American Society for Information Science and Technology*, 44(2), 101–112.
- Egghe, L. (2010). A model showing the increase in time of the average and median reference age and the decrease in time of the price index. *Scientometrics*, 82(2), 243–248.
- Egghe, L.; Ravichandra Rao, I.K. (1992). Citation age data and the obsolescence function: fits and explanations. *Information Processing & Management*, 28(2), 201–217.
- Egghe, L.; Rousseau, R. (1990). *Introduction to informetrics. Quantitative methods in library, documentation and information science*. Amsterdam, New York, Oxford, Tokyo: Elsevier Science Publishers.
- Egghe, L.; Rousseau, R. (2000). Aging, obsolescence, impact, growth and utilization: definitions and relations. *Journal of the American Society for Information Science and Technology*, 51(11), 1004–1017.
- Gargouri, Y.; Hajjem, Ch.; Larivière, V.; Gingras, Y.; Carr, L.; Brody, T.; Harnad, S. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLoS ONE*, 5(10), 1–12.
- Gilbert, N. G. (1977). Referencing as Persuasion. *Social Studies of Science*, 7, 113–122.
- Glänzel, Wolfgang (2004). Towards a Model for Diachronous and Synchronous Citation Analysis. *Scientometrics*, 60, 511–522.
- Huntington, P.; Nicholas, D.; Jamali, H. R.; Tenopir, C. (2006). Article decay in the digital environment: an analysis of usage of OhioLINK by date of publication, employing deep log methods. *Journal of the American Society for Information Science and Technology*, 57(13), 1840–1851.
- IEEE (2015). *History of IEEE* [online]. The world's largest professional association for the advancement of technology, [12.07.2015], http://www.ieee.org/about/ieee_history.html
- InCites (2014). *InCites Journal Citation Reports* [online]. Thomson Reuters, [18.09.2014], <https://jcr.incites.thomsonreuters.com/JCRHomePageAction.action>
- Kurtz, M. J.; Bollen, J. (2010). Usage bibliometrics. *Annual Review of Information Science and Technology*, 44(1), 1–64.
- Larivière, V.; Archambault, É.; Gingras, Y. (2008). Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900–2004). *Journal of the American Society for Information Science and Technology*, 59(2), 288–296.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Law, J.; French, D. (1974). Normative and interpretive sociologies of science. *The Sociological Review*, 22(4), 581–595.
- Leydesdorff, L.; Bornmann, L.; Marx, W.; Milojević, S. (2014). Referenced Publication Years Spectroscopy applied to *iMetrics: Scientometrics, Journal of Informetrics*, and a relevant subset of *JASIST. Journal of Informetrics*, 8(1), 162–174.
- Line, M.B.; Sandison, A. (1974). 'Obsolescence' and changes in the use of literature with time. *Journal of Documentation*, 30(3), 283–350.

- Liu, M. (1993). The complexities of citation practice: a review of citation studies. *Journal of Documentation*, 49(4), 370–408.
- McDonald, J. H. (2014). *Recommendation* [online]. Handbook of biological statistics, [26.07.2015], <http://www.biostathandbook.com/small.html>
- Marx, W.; Bornmann, L. (2014). Tracing the origin of a scientific legend by reference publication year spectroscopy (RPYS): The Legend of the Darwin Finches. *Scientometrics*, 99(3), 839–844.
- Marx, W.; Bornmann, L.; Barth, A.; Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751–764.
- Matricciani, E. (1991). The probability distribution of the age of references in engineering papers. *IEEE Transactions on Professional Communication*, 34(1), 7–12.
- Merton, R.K. (1973). The Normative Structure of Science. In: R.K. Merton (ed.) *The Sociology of Science. Theoretical and Empirical Investigations*. Chicago: The University of Chicago Press, 267–278.
- Motylev, V. M. (1989). The main problems of studying literature aging. *Scientometrics*, 15(1), 97–109.
- Nicholas, D.; Huntington, P.; Dobrowolski, T.; Rowlands, I.; Jamali, H. R.; Polydoratos, P. (2005). Re-visiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication. *Information Processing & Management*, 41(6), 1441–1461.
- Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science & Technology*, 41(1), 609–641.
- Norris, M.; Oppenheim, Ch.; Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59(12), 1963–1972.
- Opaliński, Ł. (2013). Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych. *Przegląd Biblioteczny*, 81(2), 152–171.
- Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, London, New York: Chapman & Hall/CRC: Taylor & Francis Group.
- Skalska-Zlat, M. (1988). Bibliometria – pojęcia, metody, kierunki badań. *Roczniki Biblioteczne*, XXXII(2), 259–283.
- Small, H. G. (2004). On the shoulders of Robert Merton. Towards a normative theory of citation. *Scientometrics*, 60(1), 71–79.
- Snarska, A. (2011). *Statystyka. Ekonometria. Prognozowanie. Ćwiczenia z Excelem 2007*. Warszawa: Wydawnictwo PLACET.
- Stefaniak, B. (1999). Cytowania literatury naukowej i ich udział w procesie komunikacji. *Zagadnienia Naukoznawstwa*, 1(139), 49–56.
- Stinson, R. E.; Lancaster, F. W. (1987). Synchronous Versus Diachronous Methods in the Measurement of Obsolescence by Citation Studies. *Journal of Information Science*, 13(2), 65–74.
- Stringer, M. J.; Sales-Pardo, M.; Nunes Amaral, L. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *Journal of the American Society for Information Science and Technology*, 61(7), 1377–1385.
- Taylor, J. R. (2012). *Wstęp do analizy błędów pomiarowych*. Warszawa: Wydaw. PWN.
- Vaughan, L. (2008). *Statistical methods for the information professional. A practical, painless approach to understanding, using, and interpreting statistics*. Medford: New Jersey: Information Today, Inc.
- Walkenbach, J. (2011). *Excel 2010 PL. Biblia*. Gliwice: Helion.
- Wallace, D. P. (1986). The relationship between journal productivity and obsolescence. *Journal of the American Society for Information Science*, 37(3), 136–145.
- Wallace, M. L.; Larivière, V.; Gingras, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, 3(4), 296–303.
- White, M. D.; Wang, P. (1997). A qualitative study of citing behavior: contributions, criteria, and metalevel documentation concerns. *The Library Quarterly: Information, Community, Policy*, 67(2), 122–154.

- WoS (2015). *Results for: Publication name: (journal of lightwave technology). Refined by: Publication years: (1988)* [online]. Web of Science Core Collection, [19.07.2015], [http://apps.webofknowledge.com/Search.do?product=UA&SID=X1eEuyaFVxYks\]lmQnY&search_mode=GeneralSearch&prID=fb2dbc34-13a7-46a3-a44f-da7589075a9b](http://apps.webofknowledge.com/Search.do?product=UA&SID=X1eEuyaFVxYks]lmQnY&search_mode=GeneralSearch&prID=fb2dbc34-13a7-46a3-a44f-da7589075a9b)
- Zhang, G.; Ding, Y.; Milojević, S. (2013). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology*, 64(7), 1490–1503.

Aneksy

- Aneks 1:* Miesiące i lata wydania źródeł cytowanych w kolejnych miesiącach roku 2014, w czterech zbadanych czasopismach z zakresu inżynierii elektronicznej [online], <http://figshare.com/s/30c51f3ca450b40ec804>
- Aneks 2:* Wyniki testu chi-kwadrat dla danych empirycznych, które zebrano w 21 przedziałach o równej szerokości, przed wykluczeniem prac starszych niż sześćdziesięcioletnie [online], <http://figshare.com/s/3572459630e62261535b>
- Aneks 3:* Wyniki testu chi-kwadrat dla danych empirycznych, które zebrano w 21 przedziałach o równej szerokości, po wykluczeniu prac starszych niż sześćdziesięcioletnie [online], <http://figshare.com/s/6316714212dae176a875>

The problem of stability of researchers' citation behaviour and the possibility of qualitative approach in the context of the methodology of ageing studies of scientific literature

Abstract

Purpose/thesis: The aim of the paper is to provide an empirical test of the hypothesis on the presence of the stability feature in researchers' citation behaviour. The stability feature can be defined as a relative constancy of behaviour during long periods of time.

Approach/methods: The authors made an attempt to validate and support a synchronous approach to the studies on scientific literature ageing. The analysis was done on the basis of a comparative study of citation behaviour of a selected group of researchers who publish in the field of electronic engineering. The authors used statistically oriented analysis to investigate the probability distribution of the age of sources cited in four scientific journals representative for the whole field. The results were compared to those from the previous study of the same journals, presented by Emilio Matricciani in 1991.

Results and conclusions: The performed statistical analysis proved that the observed probability distributions of the age of papers cited in the selected foreign journals published at two different moments in time were almost identical. It means that the same conclusion can be drawn as regards the rate of the ageing process. This outcome tentatively confirms the existence of a hypothesized feature of the citation behaviour of the researchers who publish in the field of electronic engineering.

Originality/value: The earlier literature of the field involves several papers related to the issue discussed in this paper, however only to a marginal extent. The authors who previously pointed at the need for justifying a synchronous methodology, treated this issue only as a subplot. The originality of the approach presented in this article is that (i) the problem has been clearly stated now, (ii) a statistical methodology to solve the problem has been suggested and (iii) implications crucial

for the cognitive grounds of the synchronous method in the domain of scientific literature ageing analyses have been identified.

Keywords

Bibliometrics. Scientific journals. Citation behaviour. Stability of citation behaviour. Scientific literature ageing. Statistical methods in information science.

Mgr ŁUKASZ OPALIŃSKI uzyskał tytuł magistra w zakresie nauki o języku i komunikacji (jako specjalności wyodrębnionej z kierunku filozofia) na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie w 2005 r. Ukończył studium podyplomowe „Technologie informacyjne i bibliotekoznawstwo” w Wyższej Szkole Informatyki i Zarządzania z siedzibą w Rzeszowie w 2009 r. oraz studium podyplomowe „Bibliotekarz dziedzinowy” w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych, *Przegląd Biblioteczny* 2013, 81(2), 152–171; ‘Annual aging factor’ jako narzędzie analiz starzenia się piśmiennictwa, *Przegląd Biblioteczny* 2013, 81(4), 466–484.

Kontakt z autorem:

lopa@prz.edu.pl

Biblioteka Politechniki Rzeszowskiej

al. Powstańców Warszawy 12

35–959 Rzeszów

Mgr MARCIN JAROMIN pracuje na stanowisku asystenta w Zakładzie Informatyki Chemicznej Wydziału Chemicznego Politechniki Rzeszowskiej. Tytuł magistra inżyniera uzyskał w 2004 r. na Wydziale Chemicznym Politechniki Rzeszowskiej oraz w 2005 r. na Wydziale Elektrotechniki i Informatyki Politechniki Rzeszowskiej. Specjalizuje się w dziedzinie bioinformatyki i statystyki matematycznej. Najważniejsze publikacje: Computer Aided Material Selection in Design Process. *Annals of Computer Science and Information Systems*, 2013, 1, 3–7; (współaut. B. Dębska) Komputerowo wspomagany dobór materiałów inżynierskich w procesie projektowania. W: B. Dębska, G. Fic (red.) *Systemy Informacyjne w Chemii*. Tom 3. Rzeszów: Oficyna Wydawnicza Politechniki Rzeszowskiej, 2006, 37–46; (współaut. B. Dębska, P. Peszko) Nowe metody kształcenia studentów na kierunku Technologia Chemiczna. *Przemysł Chemiczny*, 2006, 8–9, 1171–1174.

Kontakt z autorem:

mjaromin@prz.edu.pl

Zakład Informatyki Chemicznej

Wydział Chemiczny

Politechnika Rzeszowska im. I. Łukasiewicza

al. Powstańców Warszawy 6

35–959 Rzeszów

Mgr JOLANTA WIKIERA pracuje na stanowisku bibliotekarza w Pracowni Zbiorów Specjalnych i Punkcie Informacji Normalizacyjnej Biblioteki Politechniki Rzeszowskiej. Tytuł magistra bibliotekoznawstwa uzyskała w 2004 r. na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie.

Kontakt z autorką:

jw@prz.edu.pl

Biblioteka Politechniki Rzeszowskiej

al. Powstańców Warszawy 12

35–959 Rzeszów