

# Piotr Nowak

---

## WWW źródłem zmian w paradygmacie bibliometrii. Początki webometrii

---

Biblioteka 9 (18), 55-74

---

2005

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej [bazhum.muzhp.pl](http://bazhum.muzhp.pl), gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

PIOTR NOWAK

## WWW źródłem zmian w paradygmacie bibliometrii. Początki webometrii

Uwagi wstępne

Kiedy w nauce po raz pierwszy pojawiła się dojrzała refleksja teoretyczna nad ilościowymi aspektami produkcji wydawniczej, a fakt ten zwykliśmy sytuować w XX wieku pomiędzy wojnami światowymi, niewiele zjawisk zapowiadało pojawienie się elektronicznych nośników informacji. Gdy w latach 70 ubiegłego wieku zaczęto realizować pierwsze, szerzej zakrojone programy digitalizacji, nikt nie przewidywał, że ich dynamiczny rozwój realnie zagrozi źródłom tradycyjnym. Bibliometria zareagowała na zjawisko digitalizacji, włączając źródła elektroniczne do podstawy własnych badań.

Nie ma wątpliwości, że przez operowanie zbiorami dokumentów elektronicznych procedura badawcza zdecydowanie została uproszczona, łatwiej też stosować poszczególne, typowe dla pracy badawczej, schematy i techniki. Pierwsze studia nie mogły bazować jeszcze na dokumentach z sieci. Opierały się na bibliograficznych bazach danych dostarczanych off-line. Większość dokumentów elektronicznych, udostępniana właśnie w taki sposób, spełnia w istocie funkcję analogiczną do funkcji dokumentu tradycyjnego (rozumianego jako jednostka zasobu drukowanego). Dlatego też nie ma chyba specjalnego uzasadnienia traktowanie źródeł, które przyjęły postać elektroniczną przez digitalizację (tzn. zostały zapisane na nośniku cyfrowym z prymarnej postaci analogowej), jako odrębnej jakości w badaniach bibliometrycznych. Nikt z badaczy nie twierdził, że wymagają one wyjątkowej procedury metodologicznej, choć, z pewnością, pod względem technicznym prowadzi się je nieco inaczej.

Ten typ studiów, rozwijających się jeszcze dzisiaj, traktować należy jako przejściowy etap wiodący od typowych analiz bibliometrycznych do studiów *stricte* webometrycznych.

Mechanizmy rządzące światem dokumentów elektronicznych różnią się od mechanizmów funkcjonowania dokumentów tradycyjnych dopiero wówczas, gdy zmieni się model ich dostarczania. Model ten faktycznie zaczął się zmieniać szybciej, niż przypuszczano. W momencie upowszechnienia się nośnika cyfrowego nastąpiła druga rewolucja w dziedzinie źródeł informacji. Jej zasadniczym elementem było wykładnicze rozprzestrzenianie się systemów sieciowego dostarczania dokumentów. W odpowiedzi na to zjawisko pojawiła się webometria (cybermetria, netometria)<sup>1</sup>.

Najważniejsze pytanie metodologiczne, jakie nasuwa się w związku z tym faktem, brzmi: czy obecny dorobek teoretyczny bibliometrii może być – bez poddawania go poważniejszym modyfikacjom – wykorzystywany do analizy korpusu publikacji elektronicznych dostarczanych sieciowo, czy należy raczej mówić o innym niż paradygmat bibliometrii, paradygmacie uprawiania webometrii? Spróbujemy na nie odpowiedzieć po skrótowym przedstawieniu podstaw webometrii.

### Podstawowe pojęcia. Główni przedstawiciele

Powszechnie przyjmuje się, że termin *webometrics* po raz pierwszy użyty został przez T. C. Almind i P. Ingwersena w 1997<sup>2</sup>. W tym samym roku w tytule czasopisma elektronicznego pojawił się termin *cybermetrics*<sup>3</sup>. Przed rokiem 1997 sporadycznie używano terminu *bibliometria WWW*<sup>4</sup>. W tej sy-

---

<sup>1</sup> W Polsce pierwsze opracowanie wprowadzające w podstawy webometrii przygotowała M. Skalska-Zlat. Przedstawiono w nim rozwój tej subdyscypliny od chwili pojawienia się do poł. 2004 roku. Ten bardzo dobry tekst trafnie wskazuje na najważniejsze zmiany w podstawach teoretycznych bibliometrii wywołane coraz szerszym wkraczaniem Internetu w różne obszary komunikacji, w tym także oczywiście w nauce. Por. M. Skalska-Zlat, *Cybermetria, netometria, webometria – nowe pojęcia i zadania informetrii*, w: *Przestrzeń informacji i komunikacji społecznej*, Kraków 2004, s. 159-168. Jednocześnie opublikowano także dwa studia webometryczne o charakterze empirycznym: D. Kotyrasa, *Możliwości badań webometrycznych w odniesieniu do serwisów internetowych szkół wyższych*, ibidem, s. 145-148 oraz R. Sapy, *Proste badania bibliometryczne: zagraniczne odnośniki do polskich bibliotek uniwersyteckich*, ibidem, s. 154-158. To pierwsze tego typu prace w kraju, dlatego nieco uwagi poświęcimy im niżej w kontekście głównych kierunków badań webometrycznych.

<sup>2</sup> T. C. Almind, P. Ingwersen, *Informetric analyses on the World Wide Web: methodological approaches to 'Webometrics'* „Journal of Documentation”, vol. 53, nr 4, s. 404-426.

<sup>3</sup> *Cybermetrics. International Journal of Scientometrics, Informetrics and Bibliometrics*, <<http://www.cindoc.csic.es/cybermetrics.html>>, [dostęp: 13.03.2005].

<sup>4</sup> R. R. Larson, *Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace*, 1996, <<http://sherlock.berkeley.edu/asis96/asis96.html>>, [dostęp: 18.02.2005].

tuacji mechanizm powstawania terminu webometria przypominałby nieco mechanizm powstania terminu *bibliometria* (zgodnie z istniejącą w języku tendencją do holofrazy):

bibliografia statystyczna – bibliometria – bibliometria WWW  
– webometria (cybermetria/netometria etc.)

Autor nie znalazł w literaturze wcześniejszego użycia terminu *webometria*. Virgil Diodato w swym *Słowniku bibliometrii* wydanym w 1994 roku nie umieścił go na liście haseł<sup>5</sup>.

Oczywiście, webometrię definiuje się analogicznie do bibliometrii. W zasadzie mieści się ona nawet w najpopularniejszej definicji bibliometrii A. Pritcharda: „the application of mathematics and statistical methods to books and **ather media of comunication**”<sup>6</sup>.

W najbardziej lakoniczny sposób termin webometria definiowany jest chyba przez L.D. Fugl – „Webometrics is the study of quantitative aspects of webpages or nodes”<sup>7</sup>.

M. Skalska-Zlat dokonała przeglądu definicji webometrii w wybranych encyklopediach ogólnych oraz specjalnych. W *Wikipedii* znalazła następującą definicję: „The science of webometrics (also cybermetrics, web metrics) tries to measure the Internet to get knowledge about number and types of hyperlinks of the World Wide Web and usage paterns. Similar sciences are Informetrics, Bibliometrics, Scientometrics and Web Mining”<sup>8</sup>. Podobnie webometrię definiuje *Encyklopedia of Information and Library Science* (2003), a także autorzy bibliografii rejestrujących piśmiennictwo z tego zakresu<sup>9</sup>.

Za najwybitniejszych przedstawicieli webometrii uznaje się Leo Egghe, Ronalda Rousseau, Loeta Leydesdorffa i Petera Ingwersena. Sądzimy, że opieramy na przesłance, jaką jest uhonorowanie tych osób Medalem im. Dereka J. de Solla Price’a (Derek J. de Solla Price Medal). Leo Egghe i Ronald Rousseau otrzymali go w roku 2001, Loet Leydesdorff w 2003, a Peter Ingwersen i Howard D. White w 2005<sup>10</sup>.

<sup>5</sup> *Dictionary of bibliometrics*, New York: The Haworth Press 1994, s. 185.

<sup>6</sup> A. Pritchard, *Statistical bibliography or bibliometrics?*, „Journal of Documentation”, nr 25, s. 349.

<sup>7</sup> *Fundamental methodologies and tools for the employment of webometrics analyses*, <<http://dlist.sir.arizona.edu/archive/00000287>>, s. 10, [dostęp: 25.02.2005].

<sup>8</sup> *Wikipedia*, <<http://en.wikipedia.org/wiki/Webometrics>>, [dostęp: 25.02.2005]; M. Skalska-Zlat, *Cybermetria...*, op. cit., s. 161.

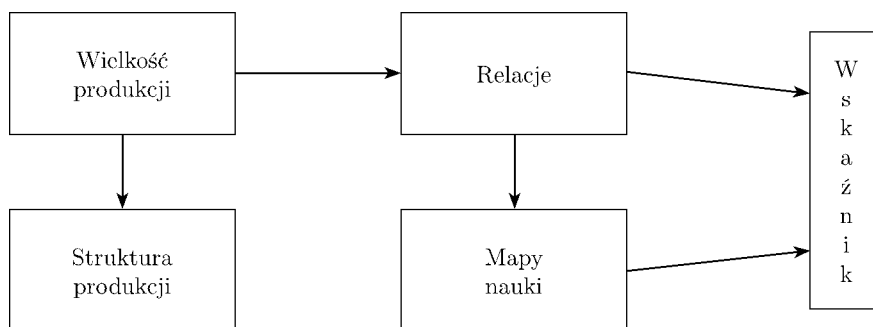
<sup>9</sup> *Ibidem*.

<sup>10</sup> *Awardees of the Derek J. de Solla Price medal*, <[http://users.pandora.be/ronald.rousseau/html/price\\_awardees.html](http://users.pandora.be/ronald.rousseau/html/price_awardees.html)>, [dostęp: 02.02.2005].

## Obszar badań

Nie będziemy dalecy od prawdy, gdy stwierdzimy, że przedmiot webometrii w zasadzie jest analogiczny do przedmiotu klasycznej bibliometrii. Różnica polega przede wszystkim na nośniku oraz sposobie dostarczania dokumentu (systemie informacyjnym, w ramach którego funkcjonuje). Teza ta, rzecz jasna, wcale nie oznacza, że dochodzimy na tym polu do analogicznych wyników czy formułujemy identyczne wskaźniki.

Z empirycznego punktu widzenia bibliometria skoncentrowana była na ustalaniu wielkości produkcji oraz charakteryzujących ją wskaźników. G. Heimeriks i P. van den Besselaar proponują pewien schemat modelu jej podstawowej metody empirycznej<sup>11</sup>. Opierając się na nim główne wymiary metodologii bibliometrii przedstawić można w taki oto sposób:



Wydaje się, że w warstwie empirycznej webometria będzie miała podobną strukturę zadań.

Do podstawowych zadań, bez których nie uda się ustalać wskaźników, należy będzie ustalenie wielkości produkcji.

## Rozmiar Internetu

Warunkiem prowadzenia studiów kwantytatywnych nad piśmiennictwem było istnienie odpowiedniej liczby dokumentów stanowiących ich podstawę. Bez dynamicznego rozwoju ruchu wydawniczego nie byłoby bibliografii statystycznej i wszystkich późniejszych metod bibliometrycznych, jakie powstały w ubiegłym wieku. Dlatego bardziej zaawansowane badania

<sup>11</sup> G. Heimeriks, P. van den Besselaar, *State of Art. In Bibliometrics and Webometrics. State of the Arts*, part B: WP8, Amsterdam: University of Amsterdam, 2002, s. 7.

kwantytatywne nad piśmiennictwem poprzedzone musiały zostać odtworzeniem rozmiarów ruchu wydawniczego. To samo dotyczy wszelkich studiów i badań webometrycznych.

Jednak w odróżnieniu od stosunkowo dokładnych statystyk książek i prasy, ustalenie rozmiarów Internetu sprawia ogromne kłopoty.

Analitycy, którzy podejmowali te próby, zgadzają się z sobą tylko w jednym punkcie: następuje szybki wzrost Internetu, gdy przychodzi jednak do konkretnych liczb są bardzo ostrożni, a dane, jakie przytaczają nie zawsze są spójne<sup>12</sup>.

S. Lawrance i C.L. Giles w grudniu 1997 roku indeksowany Web szacowali na 320 mln stron<sup>13</sup>. W rok później ci sami autorzy przeprowadzili kolejną analizę, jednak według nieco innej metodyki. Przyjęto, że w sumie istniało wówczas 16 mln serwerów. Na każdym z nich było 289 stron, co łącznie daje ok. 800 mln stron (oczywiście również bez tzw. stron głębokich<sup>14</sup>, które są poza zasięgiem zwykłych wyszukiwarek<sup>15</sup>). Widać zatem, że w zależności od przyjętego sposobu obliczenia – w praktyce prawie za ten sam okres – dane różnią się diametralnie, bo dokładnie dwuipółkrotnie.

Inkomi Cooperation i NET Research Institut szacował rozmiar Internetu na początku 2000 roku na 1 bilion dokumentów (również bez stron głębokich)<sup>16</sup>. M. Skalska-Zlat za J. Bar-Ilan cytuje jeszcze inne dane: „Szacuje się, że w 1995 roku było jedenaście milionów stron WWW, a w 2000 roku już dwa miliardy dwieście milionów. Miesięczny przyrost wynosi ok. 7 milionów stron, a liczba ta stale rośnie w zawrotnym tempie<sup>17</sup>. Podobne wielkości podaje I.F. Aquillo, który przyjmuje, że w połowie 2000 roku istniało 2 miliardy stron Web<sup>18</sup>. Przy czym, w tym samym 2000 roku szacowano

<sup>12</sup> Por. np. L.D. Fugl, *Fundamental methodologies and tools for the employment of webometrics analyses*. The Royal School of Library & Information Science, Denmark, May 2001, <<http://dlist.sir.arizona.edu/archive/00000287>>, [dostęp: 05.02.2005].

<sup>13</sup> *Searching the Worl Wide Web*, „Sciencce”, vol. 280, s. 98-100.

<sup>14</sup> *Deep web* lub *invisible web*, czyli tzw. sieć ukryta; w odróżnieniu od *surface web*, a więc sieci widocznej. To najczęściej stosowana w literaturze terminologia. Zob. np. M.K. Bergman, *The Deep web: surfacing hiddenvalue*, „Journal of Electronic Publishing”, August 2001, volume 7, issue 1, <<http://www.press.umich.edu/jep/07-01/bergman.html>> [dostęp: 05.02.2005].

<sup>15</sup> Termin z pewnością nicostry, nicmniej używany w piśmiennictwie przedmiotu (*ordinary web engines*). Zob. np. L.D. Fugt, *Fundamental...*, op. cit., s. 150.

<sup>16</sup> Ibidem, s. 50.

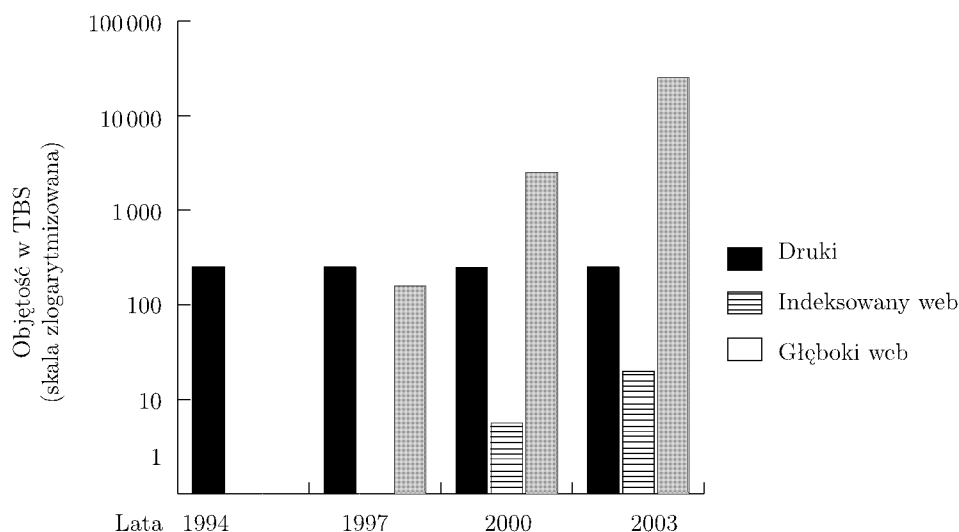
<sup>17</sup> M. Skalska-Zlat, *Cybermetria ...*, op. cit., s. 159.

<sup>18</sup> I.F. Aquillo, *Indicadores hacia una evaluacion no objectiva (cuantitativa) de sedes web*, w: *VII Jornadas Espanolas de Documentacion*. Bilbao: Servicio Editorial de la Universidad del Pais Vasco, 2000, s. 233-248. (Cyt. za: C. Faba-Percz, V.P. Guccrro-Botc, F. Moya-Ancgon, *Sitation distributions and Bradford's law in a closed Web space*, „Journal of Documentation”, vol. 59, nr 5, 2003, s. 558-580.

– Bergman – że stron głębokich jest 500 razy więcej niż indeksowanych<sup>19</sup>. Obecnie – najostrożniej szacując – mamy ponad trzy miliardy indeksowanych stron.

Bez wątpienia, w miarę dokładne dane na temat liczby stron w Internecie pozwoliłyby przede wszystkim na wiarygodne badania efektywności wyszukiwania w sieci, efektywności funkcjonowania wyszukiwarek internetowych. Bez danych na temat poszczególnych domen nie można, nawet w przybliżeniu, określić takiego wskaźnika jak kompletność wyszukiwania.

Pomimo tego rola Internetu w komunikacji globalnej rośnie nieustannie. Interesującego porównania dokonał np. M.K. Bergman. Obliczył bowiem dynamikę przyrostu informacji w dwóch rodzajach źródeł: źródłach drukowanych oraz w Internecie (z zastosowaniem podziału na tzw. strony głębokie i strony widoczne)<sup>20</sup>:



W 2002 roku ogólne zasoby Internetu w terabajtach z podziałem na kategorie kształtowały się następująco<sup>21</sup>:

Rodzaj przekazu	Rozmiar w terabajtach
Indeksowany web	167
Głęboki web	91 850

<sup>19</sup> M.K. Bergman, *The Deep Web: Surfacing Hidden Value*, USA: Bright Planet, the Internet content company, 2000, <<http://128.121.227.57/download/decpwebwhitepaper.pdf>>, [dostęp: 06.03.2005].

<sup>20</sup> Przedruk M.K. Bergman, *The Deep web....*, op. cit., [brak pag.].

<sup>21</sup> P. Lyman i in. *How Much Information?*, <<http://www.sims.berkeley.edu/research/projects/how-much-info-2003>> [dostęp: 18.03.2005].

E-mail	440 606
Pozostałe	274
<b>Ogółem:</b>	<b>532,89</b>

Warto przyjrzeć się także topografii użytkowników Internetu. W świetle tych danych, z punktu widzenia naszego quasi-podwójnego statusu obywatelskiego (w znaczeniu obywatelskim jesteśmy zarówno Polakami, jak i obywatelami Europy) nie wyglądamy źle<sup>22</sup>:

<b>Obszar</b>	<b>Liczba użytkowników</b>
Afryka	6,31 mln
Azja (Pacyfik)	187,24 mln
Europa	190,91 mln
Bliski Wschód	5,12 mln
Kanada i USA	182,67 mln
Ameryka Łacińska	33,35 mln

Bez wątplenia, ustalenie rozmiaru Internetu było jednym z warunków powstania webometrii. Ciągłe jednak poważną przeszkodą w jej rozwoju jest brak narzędzi, które pozwoliłyby precyzyjnie oszacować ilość funkcjonujących w Internecie dokumentów. W tradycyjnej bibliometrii ustalenie danych na temat skali rozwoju ruchu wydawniczego pozwoliło na bardzo precyzyjne badania nad jego strukturą. Umożliwiło także zdefiniowanie szeregu istotnych wskaźników bibliometrycznych. Brak takich możliwości w webometrii wpływa także na słabszy rozwój jej podstaw teoretycznych.

Kłopoty z ustaleniem rozmiaru Internetu nie wyczerpują listy problemów, z jakimi borykają się jego badacze, sporo ograniczeń w badaniach kwantytatywnych wynika też ze specyfiki dokumentów pojawiających się w sieci, mowa tu o takich ich właściwościach, jak semistrukturalność i heterogeniczność.

### **Dane semistrukturalne w heterogenicznym modelu źródła**

W literaturze dane semistrukturalne definiuje się jako „(...) dane, które mają nieregularną strukturę, są heterogeniczne (niejednorodne) i globalnie rozproszone. Podstawowym miejscem występowania danych semistrukturalnych są zasoby informacji w Internecie oraz zintegrowane repozytoria danych pochodzących z różnorodnych źródeł (zarówno strukturalnych, jak i semistrukturalnych)”<sup>23</sup>.

<sup>22</sup> Ibidem.

<sup>23</sup> T. Pankowski, *Modele i języki danych semistrukturalnych w integracji heterogenicznych źródeł informacji*, w: *Inteligentne metody komputerowe dla nauki, technologii i gospodarki*, pod red. Z. Kierzkowskiego, Warszawa – Poznań: SORUS, 2004, s. 181.



Problem, jaki związany jest z danymi semistrukturalnymi, jest przede wszystkim zagadnieniem czysto informatycznym, łączącym w sobie elementy baz danych, wyszukiwanie informacji oraz sztuczną inteligencję<sup>24</sup>. Niemniej nie jest on bez znaczenia także z punktu widzenia badań webometrycznych. Rzecz sprowadza się do możliwości dokładnej oceny skali i struktury wymiany informacji w wymiarze globalnym pomiędzy różnymi systemami. Jeśli problem ten nie zostanie rozwiązany, to w zasadzie nigdy nie będziemy otrzymywać w pełni wiarygodnych liczb z Internetu. Jak bowiem słusznie zauważa, cytowany wcześniej, T. Pankowski:

(...) nasza wiedza o zasobach informacyjnych tworzonych jako dane semistrukturalne, często bardzo duże i rozproszone, jest bowiem zawsze niepełna, także zapytania sformułowane są w sposób przybliżony, zatem i odpowiedzi w pewien sposób „przybliżają” odpowiedź dokładną<sup>25</sup>.

Nie sądzimy, aby ktoś zaryzykował twierdzenie, że współczesna teoria języków informacyjno-wyszukiwawczych radzi sobie z każdym problemem związanym z wyszukiwaniem dokumentów nieporównanie mniej heterogenicznych (a w zasadzie w olbrzymiej większości pod względem formalnym homogenicznych) przechowywanych w postaci skończonego zasobu drukowanego w bibliotekach. Jak w tej sytuacji uzyskiwać w miarę pełną kompletność i dokładność wyszukiwania w Internecie?

Problem ten nie doczekał się jeszcze bardziej jednoznacznych rozstrzygnięć i na gruncie rysującego się obszaru webometrii nie był chyba dotąd szerzej przez nikogo dyskutowany. Dlatego, rzecz jasna, nie będziemy formułować na siłę własnych konkluzji. Jedno wydaje się jednak pewne: we wszelkich kwantytatywnych badaniach dokumentów dostarczanych poprzez sieć należy pamiętać o ograniczeniach, które rodzi wyraźna semistrukturalność i heterogeniczność analizowanych źródeł.

### **Dokument dynamiczny jako jednostka obliczeniowa w webometrii**

Poza wspomnianą semistrukturalnością oraz heterogenicznością źródeł wirtualnych w literaturze pojawiło się pojęcie dokumentu dynamicznego<sup>26</sup>. W związku z coraz większym zakresem operacji wykonywanych przez

---

<sup>24</sup> Ibidem, s. 181.

<sup>25</sup> Ibidem, s. 182.

<sup>26</sup> Z. Kierzkowski, A. Kołopiończyk, *Dokumenty dynamiczne w budowie kooperacyjnych systemów informatycznych*, w: *Inteligentne metody komputerowe...*, op. cit., s. 131.

współczesne systemy informacyjne, dość istotnie zmienia się i zmieniał się będzie nadal model architektury systemu informacyjnego<sup>27</sup>, którego podstawą staną się dokumenty mające zdolność głębokiej ewolucji struktury w zależności od funkcji, jaką w danym momencie będą pełnić. W bardzo krótkim czasie zmieniać się będzie także język opisu takich dokumentów oraz – siłą rzeczy – ich liczba. Ostatni z wymienionych punktów z pewnością wpłynie na paradygmat webometrii, gdyż w sposób ciągły zmianie w czasie ulegać będzie liczba dokumentów stanowiących podstawę analiz.

W tej sytuacji pojawienie się dokumentu dynamicznego może poważnie skomplikować definiowanie podstawowej jednostki obliczeniowej. Operować będziemy nie, jak dotychczas, dokumentami czy źródłami, lecz pojęciem bliższym jednostkowej informacji. Będzie to jakaś hipotetyczna faseta informacji. W chwili obecnej w zasadzie nie znajdujemy w prowadzonych badaniach webometrycznych ujęć, w których szeroko dyskutowano by pojęcie dokumentu dynamicznego. Problem ten sygnalizuje M. Skalska-Zlat, powołując się na studia L. Egghe<sup>28</sup>.

### Rozwój teorii. Badania relacji

Interesujące studia nad metodyką prowadzenia badań metrycznych nad cyberprzestrzenią wyszły spod pióra T.C. Almindy, który w 1997 roku wspólnie z P. Ingwersenem opublikował, cytowany już wcześniej, artykuł *Informetric analyses on the world wide web: methodological approaches to 'Webometrics'*<sup>29</sup>.

J. Kim przyjrzał się niektórym metodom oceny wartości materiałów naukowych publikowanych w sieci. Efektem tych badań był m.in. artykuł *Motivations for hyperlinking in scholarly electronic articles: a quantitative study*<sup>30</sup>. Poza tym wkład w rozwój badań webometrycznych wnieśli tacy badacze, jak S. Lawrence oraz C.L. Giles<sup>31</sup>. Aktywnym analitykiem w dziedzinie webometrii jest M. Thelwall<sup>32</sup>.

---

<sup>27</sup> Ibidem, s. 131.

<sup>28</sup> M. Skalska-Zlat, *Cybermetria...*, op. cit., s. 163; L. Egghe, *New informetric aspects of the Internet: some reflections – many problems*, „Journal of Information Science”, vol. 26, nr 5, s. 329-335.

<sup>29</sup> „Journal of Documentation”, vol. 53, nr 4, s. 404-426.

<sup>30</sup> „JASIS”, vol. 5, nr 10, s. 887-899.

<sup>31</sup> *Searching the Word Wide Web*, „Science”, vol. 280, s. 98-100; idem, *Accessibility of information on the web*, „Nature”, vol. 400, s. 107-109.

<sup>32</sup> *What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation*, <<http://informationr.net/ir/8-3/paper151.html>> [dostęp: 02.25.2005].

Przeglądowe studia podstaw metodologicznych webometrii wyszły spod pióra L. D. Fugl (wspomniane już dzieło *Fundamental methodologies and tools for the employment of webometrics analyses*) oraz G. Heimeriksa i P. van den Besselaara (*State of the Art. In Bibliometrics and Webometrics*)<sup>33</sup>.

Inni znani przedstawiciele webometrii to A. Woodruff, który jako jeden z pierwszych zajmował się strukturą dokumentów publikowanych w ramach WWW<sup>34</sup>, R. Larson tworzący mapy nauki w oparciu o analizę WWW<sup>35</sup> czy A.G. Smith, skupiony m.in. na empirycznym wykorzystywaniu wskaźników cytowań w cyberprzestrzeni oraz definiowaniu różnic w zasadach cytowań pomiędzy Worl Wide Web a zasobem drukowanym<sup>36</sup>.

Nieco szerzej omówić wypada pierwsze polskie empiryczne dokonania na polu webometrii<sup>37</sup>. D. Kotyras przeprowadził pilotażowe badania nad powiązaniem pomiędzy stronami Web polskich uniwersytetów. Wcześniej podobne analizy wykonali wspomniani już M. Thewall i A. Smith. Ten ostatni sam zresztą też prowadził takie badania, analizując wystąpienia różnych typów domen; m.in. geograficznych, np. as, nz, czy sektorowych, np. as.com<sup>38</sup>.

Celem studiów D. Kotyrasa było obliczenie dla stron 17 polskich uczelni liczby wzajemnych odesłań. Ustalił on, że najczęściej wychodzących linków mają Uniwersytety: Warszawski (131), Szczeciński (80), UAM (73) oraz Wrocławski (62). Z kolei pod względem odesłań przodują UW (62), Uniwersytet Gdański (50), UAM (47) oraz Jagielloński i Śląski (po 46)<sup>39</sup>.

R. Sapa poszukiwał natomiast zagranicznych odnośników do stron WWW czterech dużych polskich bibliotek akademickich:

<sup>33</sup> Universiteit van Amsterdam, January 2002.

<sup>34</sup> Por. A. Woodruff et al., *An investigation of documents from the World Wide Web. Proceedings of the Fifth International World Wide Web Conferens*. May 6-10, 1996, Paris, France 1996, <[http://www5conf.inria.fr/fich\\_html/papers/P7/Overview.html](http://www5conf.inria.fr/fich_html/papers/P7/Overview.html)> [dostęp: 06.03.2005].

<sup>35</sup> *Bibliometric of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspaces*, 1996, <<http://sherlock.berkeley.edu/asis96/asis/96.html>>, [dostęp: 06.03.2005].

<sup>36</sup> M.in. *Web links as analogues of citation*, „Information Research”, vol. 9, no 4, July 2004, <<http://informationr.net/ir/9-4/paper188html>> [dostęp: 2005-03-06].

<sup>37</sup> Przy okazji warto zaznaczyć, że od chwili użycia w światowej literaturze terminu webometria do momentu jego pojawienia się w polskim piśmiennictwie minęło 7 lat (1977-2004). Od użycia w 1969 roku przez A. Pritcharda terminu bibliometria do momentu pierwszego pojawienia się go w polskiej monografii Krzysztofa Migonia w 1984 roku minęło 15 lat. Może to świadczyć o szybszym, niż w przeszłości odrabianiu przez polską naukę dystansu do nauki światowej.

<sup>38</sup> A. Smith, *ANZAC webometrics: exploring Australasian Web structures*, 1999, <<http://www.csu.edu.au/special/online99/proceedings99/203b.htm>>, [dostęp: 11.02.2005].

<sup>39</sup> D. Kotyras, *Możliwości badań...*, op. cit., s. 146-148.

- Biblioteki Uniwersytetu Warszawskiego,
- Biblioteki Uniwersytetu Jagiellońskiego,
- Biblioteki Uniwersytetu im. A. Mickiewicza w Poznaniu,
- Biblioteki Uniwersytetu Wrocławskiego.

Okazuje się, że największa liczba linków do wymienionych bibliotek biegnie z niemieckiego Web (32), a następnie: amerykańskiego (27), włoskiego (13), organizacji międzynarodowych (9), kanadyjskiego (6), rosyjskiego (6) oraz brytyjskiego (6)<sup>40</sup>.

W interpretacji wyników tych analiz obaj autorzy ograniczeni zostali typowymi dla obecnych badań webometrycznych problemami heurystycznymi. Po pierwsze, przy niewielkich różnicach wyników (a liczby linków biegnących do polskich bibliotek uniwersyteckich w badaniach R. Sapy różnią się nieznacznie), analizy przeprowadzone w innym czasie mogą przynieść istotnie różniące się rezultaty. W wyniku zreindeksowania materiału nawet przez tę samą wyszukiwarkę otrzymamy odmienne dane.

Po drugie, kontrowersje budzi przyjmowanie liczby odnośników do strony Web jako miary wartości tej strony. A zatem, nawet gdyby udało się otrzymać niezmiennie w czasie wyniki, to i tak na obecnym etapie refleksji teoretycznej webometrii ich interpretacja może nasuwać wątpliwości. Problemy te oczywiście mają charakter uniwersalny dla wszystkich badań tego typu, a nie tylko wspomnianych dwóch polskich analiz.

### **Web Impact Factor** – przykład wskaźnika webometrycznego

Badania środowiska Web zaowocowały oryginalnym, typowym dla webometrii wskaźnikiem. Jest nim tzw. Web Impact Factor. W tej chwili relacje w cyberprzestrzeni nie mają jeszcze takiej wagi, jak w przypadku źródeł drukowanych, głównie dlatego, że występujące tam dokumenty odgrywają nieco inną rolę w komunikacji w nauce. W kategoriach wartościujących mają zdecydowanie mniejsze znaczenie. Problem w dużym stopniu sprowadza się do nieokreślonego jeszcze dokładnie miejsca w komunikacji naukowej czasopism elektronicznych. Słaby rozwój czasopism elektronicznych, które (jak niektórzy sądzą) miały szybko wyeliminować czasopisma drukowane, z pewnością zaskoczył analityków i rozczarował użytkowników. Wprawdzie w drugiej połowie lat 90 notowano dość szybki wzrost tytułów czasopism dostępnych w sieci, niemniej najczęściej były to odpowiedniki wersji drukowanych. M. i M. Nahotko przedstawili strukturę elektronicznego zasobu

---

<sup>40</sup> R. Sapa, *Proste badania...*, op. cit., s. 156.

czasopism w Bibliotece Uniwersyteckiej w Regensburgu. Jak się okazuje, w 2003 roku tylko 1 tytuł na 10 był tzw. rzeczywistym czasopismem elektronicznym (nie miał odpowiednika drukowanego)<sup>41</sup>. Dlatego obecny zasób akademickich czasopism tego typu, jeśli nie mówimy o odpowiednikach tytułów drukowanych, nie spełnia funkcji, jakimi obarczamy najlepsze czasopisma tradycyjne. Poza tym, w źródłach w sieci cytowany jest także zasób drukowany (oraz odwrotnie), a opracowywane na potrzeby webometrii wskaźniki dotyczą wyłącznie relacji pomiędzy dokumentami sieciowymi, bowiem są obliczane na podstawie linków hipertekstowych.

W założeniu Web Impact Factor jest wskaźnikiem, który ma spełniać podobną rolę do wskaźnika wpływu stosowanego przez Instytut Informacji Naukowej w Filadelfii<sup>42</sup>. WIF zaproponował Ingwersen i nie ma kontrowersji w metodyce jego obliczania, niektórzy badacze stosują go wyłącznie do oceny czasopism naukowych, inni oceniają przy jego pomocy również strony nie mające nic wspólnego z prasą elektroniczną.

WIF jako wskaźnik wyraźnie inspirowany wskaźnikiem wpływu ma odzwierciedlać miarę zainteresowania daną stroną. Ścisłej: jest on stosunkiem liczby odnośników (linków) do określonego czasopisma elektronicznego (lub innego dokumentu w sieci) do liczby stron WWW tego czasopisma (lub innego dokumentu w sieci). Czas nie odgrywa w metodyce jego obliczania żadnej roli. Przy czym obie wielkości powinny być generowane za pomocą tej samej wyszukiwarki.

W tej sytuacji, jeśli liczbę linków do określonego czasopisma oznaczymy symbolem  $L$ , natomiast liczbę stron WWW tego czasopisma jako  $W$ , otrzymamy prostą formułę obliczania Web Impact Factor:

$$WIF = \frac{L}{W}$$

Badania empiryczne z wykorzystaniem WIF nie są prowadzone do tego stopnia celowo, jak obliczanie IF czasopism indeksowanych przez ISI, co w praktyce oznacza, że zaledwie eksperymentalnie ustalono WIF dla niektórych wybranych tytułów czasopism elektronicznych. U podstaw tych

<sup>41</sup> M. i M. Nahotko, *Powstanie i rozwój czasopism elektronicznych*, w: *Przestrzeń informacji i komunikacji społecznej*, pod red. M. Kocójowej, 2004, s. 127. W związku z szybkim tempem przyrostu elektronicznych odpowiedników wersji drukowanych, przy niewielkiej progresji czasopism elektronicznych nie będących takimi odpowiednikami już w 1994 roku G. van Marle sugerował nazywanie tych ostatnich *real electronic journal*. Zob. tegoż *Electronic Serial Publishing and its effect on the traditional information chain*, „Serials”, vol. 7, nr 1 (March), s. 17-28.

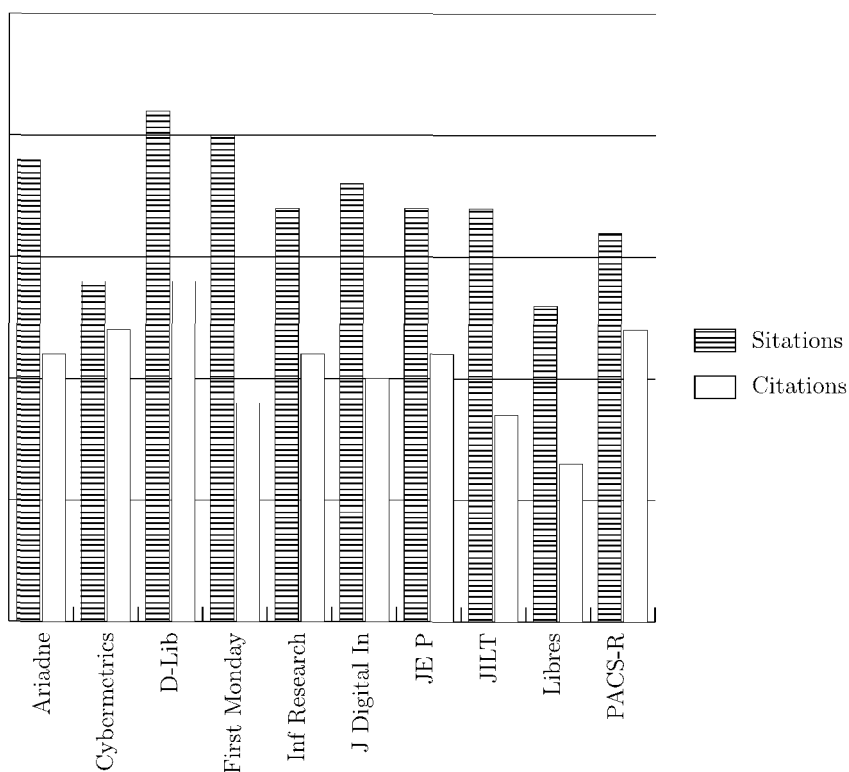
<sup>42</sup> Obliczanie wskaźnika wpływu (Impact Factor) przedstawiono na stronie: *The Impact Factor*, <<http://www.isinet.com/essays/journalcitationreports/7.html>>, [dostęp: 31.03.2005].

prac leży oczywiście analogiczna idea postępowania, ale tworzenie linków do stron, czyli owe cyberprzestrzenne cytowania zwykło określać się mianem sytowań (*sitation*).

Alistar G. Smith<sup>43</sup> porównał cytowania obliczane przez ISI do sytowań webometrycznych uzyskiwanych przez następujące czasopisma:

- „Ariadne”,
- „Cybermetrics”,
- „D-Lib Magazine”,
- „First Monday”,
- „Information Research”,
- „Journal of Digital Information”,
- „Journal of Electronic Publishing”,
- „Journal of Information, Law and Technology”,
- „Libres: Library and Information Science Research Electronic Journal”,
- „PACS-R: Public Access Computer Systems Review”.

Uzyskane przez wymienione czasopisma cytowania ISI oraz sytowania (Alta Vista) przedstawione zostały na poniższym wykresie:



<sup>43</sup> A. G. Smith, *Citations and...*, op. cit., [brak pag.].

Powyższe dane zachęcają do kilku, co najmniej, uwag. Po pierwsze, gołym okiem widać niewielką korelację pomiędzy wskaźnikami, jakimi są cytowania i sytowania. Potwierdza to oczywiście Smith, który przeprowadza następnie porównanie Web Impact Factor z notowanymi dla tych czasopism przez ISI cytowaniami. W tym wypadku korelacja jest nieco większa, jednak w dalszym ciągu daleko jej do istotności. W tej sytuacji należy postawić pytanie, co w zasadzie odzwierciedla WIF? M. Skalska-Zlat sugeruje, że WIF mógłby być użytecznym uzupełnieniem badań prowadzonych za pomocą tradycyjnego IF – porównania wskazywałyby status stron WWW. Może także pełnić rolę wskaźnika atrakcyjności krajów, badań naukowych i instytucji<sup>44</sup>.

Problem należy jednak do bardziej złożonych, jeśli bowiem, choćby częściowo, traktować WIF w kategoriach IF, to powinien on wyraźniej oscylować ku mierze wartości. Nie będziemy chyba dalecy od prawdy, jeśli stwierdzimy, że WIF w istocie jest tylko próbą poszukiwania sposobów oceny stron WWW. Dodatkowo jest mało wyrazisty, ponieważ mówi się o jego trzech – istotnie się różniących – kategoriach:

1. **the self-links WIF**, odpowiadający pojęciu samocytowania (biorący pod uwagę linki znajdujące się w danym dokumencie do innych stron tego samego dokumentu);
2. **the external WIF**, biorący pod uwagę linki do danej strony Web od stron zewnętrznych;
3. **the overall WIF**, obliczany na podstawie wszystkich linków do danej strony (zarówno wewnętrznych, jak i zewnętrznych)<sup>45</sup>.

Jak powiedzieliśmy już wyżej, wielu autorów stosuje Web Impact Factor nie tylko w odniesieniu do czasopism, ale także do domen, instytucji (np. uczelni wyższych) i innych stron, które nie są czasopismami typu „e”. Smith na przykład obliczył WIF dla domen krajów Południowo-Wschodniej Azji, stron www wszystkich uniwersytetów Australii i Nowej Zelandii, czasopism elektronicznych ukazujących się w tych dwóch krajach i stron dwóch bibliotek Narodowych: Australii i Nowej Zelandii<sup>46</sup>.

Podsumowując dotychczasową dyskusję nad Web Impact Factor, widzimy wyraźnie, że rola, jaką ma pełnić nie została jeszcze precyzyjnie określona, stąd też wyniki wszelkich badań empirycznych prowadzonych z użyciem tego wskaźnika nie pozwalają jeszcze na budowanie poważniejszych uogólnień. Przede wszystkim nie widać w literaturze wspólnego stanowiska dotyczącego motywów tworzenia linków do stron WWW. Próby stworzenia z WIF analogicznego narzędzia do klasycznego Impact Factor, czy dopasowy-

<sup>44</sup> M. Skalska-Zlat, *Cybermetria...*, op. cit., s. 164.

<sup>45</sup> T.C. Almind, P. Ingwersen, *Informetric analyses...*, op. cit.

<sup>46</sup> A.G. Smith, *Citations and .....*, op. cit., [brak pag.].

wanie do, tworzonych w Internecie linków tych wskaźników, które z powodzeniem stosuje się w procesie oceny czasopism drukowanych bardziej chyba przeszkadza, niż pomaga rozwojowi oryginalnej myśli teoretycznej nowej dziedziny wiedzy, jaką jest webometria. Mechanizmy publikowania w Internecie (poza czasopismami „e”) różnią się tak istotnie od tradycyjnych zasad publikowania, że adaptowanie bibliometrycznych narzędzi opisu do opisu i oceny Internetu nie zawsze jest drogą właściwą. Jednak prawdopodobnie jesteśmy skazani na pewien okres przejściowy, w którym takie adaptacje będą miały miejsce. Skończy się to, gdy rola Internetu w komunikacji w nauce stanie się równie ważna, jak obecna rola systemów tradycyjnych. Warunkiem tych zmian będzie rozbudowa kontrolowanego Internetu, np. *subject gateways*. Pod tym pojęciem kryje się uporządkowany tematycznie obszar Internetu, który zawiera wyselekcjonowane pod względem poziomu merytorycznego, przejrzane przez specjalistów i dość precyzyjnie skatalogowane dokumenty, najczęściej o charakterze naukowym lub fachowym. Serwisy tego typu zamieszczają linki do dokumentów zapisanych zwykle w formacie HTML. Liczą od ok. 2 tys. do 10 tys. każdy. „Subject gateways” stanowią obecnie margines zasobów sieciowych. Wiele wskazuje jednak na to, że serwisy tego typu będą się rozwijać<sup>47</sup>.

### **System praw w webometrii: status quo czy konieczność rewizji**

To zagadnienie traktować musimy jako jeden z najistotniejszych problemów dyskusji nad ewolucją paradygmatu bibliometrii w związku z rozległym obszarem nowych analiz opartych na zbiorach elektronicznych.

Przyjmując, że jądrem teorii danej dyscypliny jest system sądów, zwanych najczęściej prawami nauki (w bibliometrii funkcjonują tzw. statystyczne prawa nauki<sup>48</sup>), należy przede wszystkim przyjrzeć się ewentualnej ewolucji tych praw na gruncie webometrii. Będziemy mieć w tym miejscu do czynienia z ze wszech miar interesującymi zagadnieniami. Skrótowno można je zawrzeć w pytaniu: czy w dyscyplinie, w której podstawą badań są systemy dokumentów elektronicznych, dojdzie do modyfikacji istniejących praw, albo – pojawienia się nowych praw, uwzględniających specyfikę zjawisk w Internecie.

Dotychczasowe prawa bibliometryczne, zarówno te najbardziej znane (Bradforda, Lotki oraz Zipfa)<sup>49</sup>, jak i mniej znane (Bootha, Brooksa, Estroupa,

<sup>47</sup> Szerzej zob.: L. Derfert-Wolf, *Serwisy tematyczne o kontrolowanej jakości w Internecie – subject gateways*, „Ebib”, nr 6/2004 (tam obszerna literatura).

<sup>48</sup> W. Krajewski, *Prawa nauki. Przegląd zagadnień metodologicznych i filozoficznych*, Warszawa: Książka i Wiedza, 1998.

<sup>49</sup> V. Diodato, *Dictionary...* op. cit., passim.



Leimkuhlersa, Price'a czy Willisa)<sup>50</sup>, dotyczą albo mechanizmów rozkładu pewnych cech piśmiennictwa (przede wszystkim prawo rozproszenia Bradforda i prawo produktywności naukowej Lotki), albo pewnych cech tekstu (jak wiadomo niektóre aspekty językoznawstwa statystycznego traktowane są również jako część obszaru bibliometrii).

Zacznijmy od przedyskutowania drugiej grupy praw, ponieważ tu sytuacja wydaje się najprostsza. Otóż dokumenty elektroniczne opierają się na przekazach tekstowych bez względu na formę ich dostarczenia i nie wydaje się, aby przekazy tekstowe nowego typu, czyli dokumenty elektroniczne – obojętnie, czy będą to zdigitalizowane zasoby drukowane czy prymarne dokumenty sieciowe – różniły się od tekstu archiwizowanego na nośniku analogowym. Nowy obszar badań pojawi się w momencie analizy np. aspektów efektywności dostarczania dokumentów w kontekście językowych konstrukcji ich wyszukiwania (choćby wykrywania relacji np. pomiędzy metodą konstruowania zapytania w języku naturalnym a skutecznością wyszukiwania dokumentu). Skądinąd w obszarze badań mieszczących się nawet w szerokich ramach nauki o informacji sporadycznie prowadzono studia empiryczne nad tekstem. Pomimo dość intensywnego rozwoju teorii języków informacyjno-wyszukiwawczych, zwłaszcza w drugiej połowie ubiegłego wieku, prawo Zipfa rzadko pojawiało się w orbicie zainteresowań nauki o informacji. W przypadku webometrii próbę jego wykorzystania do analizy popularności stron podjął tylko Nielsen<sup>51</sup>.

Nieco inaczej rzecz ma się z prawami dotyczącymi rozkładów cech piśmiennictwa. Badania takie były podejmowane, ale bazowano w nich na istniejącej teorii. Można by zatem twierdzić, że obecnie jesteśmy na etapie testowania praw bibliometrycznych na danych, które dostarcza webometria. Nie będziemy chyba dalecy od prawdy, przyjmując, że czym więcej takich badań, tym bliżej odkrycia nowych praw i prawidłowości odnoszących się tym razem do środowiska Web. Jak powiedzieliśmy wcześniej, nurt tego typu studiów nie jest jeszcze zbyt szeroki, dlatego bez większego trudu można przedstawić najważniejsze dokonania na tym polu.

W 1997 roku J. Bar-Ilan opublikował w „Scientometrics” artykuł, w którym przedstawił rozkład publikacji w sieci poświęconych tzw. chorobie wściekłych krów (BSE)<sup>52</sup>. Jednak analiza ta nie wykazała, że istniejący rozkład przypomina rozkład Bradforda<sup>53</sup>.

---

<sup>50</sup> Ibidem, *passim*.

<sup>51</sup> Za: M. Skalską-Zlat, *Cebermetria...*, op. cit., s. 167.

<sup>52</sup> J. Bar-Ilan, *The mad cow disease, Usenet newsgroups and bibliometric laws*, „Scientometrics”, vol. 39 no 1, s. 29-55.

<sup>53</sup> Jak wiemy, prawo pierwotnie sformułowane przez S.C. Bradforda dotyczyło produktywności czasopism, a konkretnie rozkładu artykułów na dany temat w określonej grupie

W dwa lata później (1999) L. Cui przeprowadziła badania rozkładu linków wychodzących od stron największych szkół medycznych w USA, także poszukując korelacji z rozkładem Bradforda. Tym razem udało się dowieść tzw. strefy jądra oraz charakterystycznie przyrastających pod względem wielkości pozostałych stref (zgodnie z prawem Bradforda wzrost ten opisywany jest formułą  $1:n:n^2\dots$ )<sup>54</sup>.

Natomiast w 1997 roku R. Rousseau badał rozkłady sytowań w Internecie. Tym razem za punkt wyjścia przyjął prawo Lotki, ekstrapolując rozkład produktywności autorskiej na rozkład sytowań do strony. Przyjęta hipoteza zakładała, że będziemy mieć do czynienia z zasadą odwrotności kwadratu Price'a (*inverse square law*). Autor tych badań uznał ostatecznie, że prawdopodobieństwo, iż dokument w sieci otrzyma „i” sytowań, statystycznie wynosi:

$$\frac{1}{i^X}$$

Przy czym  $X > 1$ . Formułę tę próbował weryfikować Broder i in., ustalając, że  $X = 2,1$ <sup>55</sup>. Wynik ten przybliżył nas do obliczeń Lotki, który  $X$  szacował na 1,888 (dla chemii) oraz 2,02 (dla fizyki). Jak wiemy, Price nieco zmodyfikował to prawo, twierdząc, że:

[...] liczba autorów  $n$  prac jest proporcjonalna do  $1/n^2$ , czyli, że produktywność jest odwrotnie proporcjonalna do kwadratu liczby tych prac. Na każdych 100 autorów jednej rozprawy przypada 25 autorów 2 prac, 11 autorów 3 pozycji itp.<sup>56</sup>

Zgodnie zatem z hipotezą zespołu kierowanego przez Brodera na każdych 100 dokumentów w Internecie, które otrzymały jedno sytowanie, przypadać będzie 25 dokumentów z dwoma sytowaniami, 11 z trzema itd.

Interesującą i stosunkowo szeroką analizę, której celem było poszukiwanie korelacji tym razem z oboma rozkładami naraz, Lotki i Bradforda, przeprowadził zespół trzech autorów: C. Faba-Perez, V.P. Guerrero-Bote

---

czasopism. W badaniach webometrycznych często ekstrapoluje się je na linki poświęcone określonymu tematowi, strony mając taką, a nie inną domcnę etc. Szerzej nt. prawa rozproszenia Bradforda zob. R.M. Lossce, *The science of information*, San Diego 1990, a także <<http://www.lis.uiuc.edu/~jdownie/biblio/bradford.htm>>, [dostęp: 02.04.2004].

<sup>54</sup> L. Cui, *Rathing health web sites using the principles of citation analysis: a bibliometric approach*, „Journal of Medical Research”, vol. 1, nr e4, <[www.jmir.org/1999/1/eh/index.html](http://www.jmir.org/1999/1/eh/index.html)>, [dostęp: 31.03.2005].

<sup>55</sup> A. Broder i in., *Graph structure in the Web*, „Computers Networks and ISDN Systems”, vol. 33, nr 1-6, s. 309-320.

<sup>56</sup> D.J. de Solla Price, *Mała Nauka – Wielka Nauka*, Warszawa: PWN, 1967, s. 46-47.

oraz F. Moya-Augenon<sup>57</sup>. Ich badania mieszczą się w typowym *case studies* i dotyczą rozkładów hiperlinków otrzymywanych przez strony poświęcone regionowi Ekstremadura położonemu w zachodniej części Hiszpanii. W zależności od rodzaju strony parametr  $X$  wahał się od 2,8 do 1,5. Dlatego trudno w tym wypadku mówić o potwierdzeniu rozkładów Lotki. W rozkładzie hiperlinków do tych stron nie potwierdzono także prawa Bradforda.

### Narzędzia bibliometrii Web

Tradycyjna bibliometria nie może się już rozwijać bez olbrzymiego w tej chwili instrumentarium, często nazywanego przemysłem bibliometrycznym<sup>58</sup>. Rozwój badań webometrycznych, rzecz jasna, jeszcze bardziej zależeć będzie od narzędzi technicznych pozostających w jej dyspozycji. Dalszy rozwój Internetu oraz pogłębianie się semistrukturalności i heterogeniczności dokumentów spowoduje konieczność przeznaczania olbrzymich sum na ten cel. Najbardziej rozbudowany obecnie sektor przemysłu bibliometrycznego, jakim z pewnością są indeksujące bazy danych, przypuszczalnie zmieni się diametralnie, bowiem w sensie technicznym nie da się zasad dotychczasowego indeksowania czasopism stosować w odniesieniu do czasopism typu „e”, obojętnie, czy wykorzystywać będziemy Web Impact Factor czy inny wskaźnik.

Jednak w chwili obecnej nie istnieją jeszcze żadne, stworzone wyłącznie do celów webometrycznych, techniczne narzędzia analiz. W praktyce badawczej wykorzystuje się zwykle wyszukiwarki internetowe. Przy czym prawie wszyscy podkreślają ich niedoskonałość w przypadku prób jakichkolwiek precyzyjniejszych badań.

Znakomita większość studiów webometrycznych przeprowadzona została za pomocą wyszukiwarki Alta Vista, albowiem tylko ona pozwala na stosowanie w odniesieniu do całego Internetu algebry Boole'a<sup>59</sup>. Rekonstrukcja zapytań prowadzi, niestety, do zmian wyników wyszukiwania. Aby np. znaleźć zewnętrzne linki do strony czasopisma elektronicznego, które posiada URL **xxx**, operuje się prostą komendą: **link: xxx and not host: xxx**. Przetawienie szyku komendy względem operatora logicznego może dać odmienny wynik.

Niektórzy autorzy sugerują korzystanie z kilku wyszukiwarek jednocześnie, jednak w praktyce rzadko jest to możliwe, bowiem – jak powie-

---

<sup>57</sup> C. Faba-Perez, V.P. Guerrero-Bote, F. Moya-Augenon, *Sitation distributions...*, op. cit.

<sup>58</sup> I. Marszakowa-Szajkiewicz, [bez tyt.], <<http://www.forumakad.pl/archiwum/2000/06/artykuły/12-polemiki.htm>>, [dostęp: 31.03.2005].

<sup>59</sup> <<http://www.altavista.com/help/search/syntax>>, [dostęp: 31.03.2005].

dzieliśmy wyżej – tylko Alta Vista pozwala na globalne przeszukiwanie Interenetu z wykorzystaniem algebry Boole'a. Niewiele także nasze badania zyskają na wiarygodności, gdy będziemy je powtarzać kilkakrotnie w stałych odstępach czasu, ponieważ prawdopodobieństwo umocnienia błędu jest w przypadku takiego zabiegu mniej więcej takie samo, jak jego neutralizacja. (Wielu badaczy intuicyjnie sugeruje taką operację, aby zminimalizować fałszywe wyniki).

Semistrukturalność i heterogeniczność Internetu, a zwłaszcza wspomiane już zjawisko dynamiki dokumentów internetowych rodzi także problemy w relacjach z czasem. Szczególną uwagę zwraca na to M. Skalska-Zlat:

Różnie jest [...] rozumiana data strony (gdy chcemy zastosować ograniczenie czasowe). Np. Alta Vista przez czas rozumie datę ostatniej aktualizacji, Northern Light w ogóle nie wyjaśnia, co rozumie przez datę. Zdarzają się przypadki podawania daty utworzenia strony lub jej modyfikacji albo nawet ostatniego wyszukania. Jest też problem z ok. 10% dokumentów nie zawierających w ogóle daty lub oznaczonych datą fałszywą, np. taką, która jeszcze nie wystąpiła [...] <sup>60</sup>.

### Dalszy rozwój teorii

Dalszy rozwój wirtualnych systemów informacyjnych wprawdzie w dużym stopniu zależy od trudno przewidywalnych zachowań komercyjnego sektora dostawców poszczególnych składników architektury tych systemów, jednak nie wydaje się możliwe, aby stracił on swoją dotychczasową dynamikę. Można przewidywać raczej przyspieszenie tego procesu. W tej sytuacji pojawi się naturalna cezura zamykająca obecny okres teorii bibliometrii trwający w zależności od definiowania kryteriów tej metody od ok. 80-120 lat. Jeśli faktycznie nastąpi naturalne niejako odejście od dotychczasowej teorii, to z pewnością nie będzie to rewolucja, bowiem nie wydaje się, aby pewne jej elementy, przede wszystkim prawa, nie miały zastosowania w odniesieniu do zbiorów dokumentów wirtualnych (wprawdzie o dość specyficznych atrybutach, jakimi są semistrukturalność i heterogeniczność), niemniej w sensie funkcjonalnym w dużym stopniu analogicznych do obecnych źródeł. Rewolucja może nastąpić w momencie szerszego pojawienia się wspomnianego dokumentu dynamicznego, albowiem wówczas zmianie – i to z pewnością istotnej – ulegną jego funkcje. Będą to dokumenty zarządzające wiedzą, a nie tylko dostarczające dane. Wówczas może nastąpić

---

<sup>60</sup> M. Skalska-Zlat, *Cybermetria...*, op. cit., s. 165.

druga, właściwa w znaczeniu metodologicznym, faza rozwoju webometrii, jeśli traktować będziemy ją wąsko, tzn. jako metodę. Wtedy także wielu dotychczasowych zjawisk na tym gruncie nie będzie można już wyjaśnić istniejącymi prawami. Gdyby ten scenariusz się sprawdził, to obecną fazę rozwoju bibliometrii można traktować jako fazę przejściową pomiędzy paradygmatem dwudziestowiecznym a nową teorią webometrii<sup>61</sup>. Trudno, co oczywiste, wyrokować, kiedy to nastąpi, ale osiągnięcia fazy przejściowej, w jakiej uczestniczymy w chwili obecnej, mogą ją naturalnie przyspieszyć. To pierwszy z czynników kształtujących ten proces. Drugim będą z pewnością zmiany w Internecie.

---

<sup>61</sup> Dokładnie rzecz ujmując, faza ta rozpoczęła się już w ubiegłym wieku. Por. np. R.R Larson, *Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace*, 1996, <<http://sherlock.berkeley.edu/asis96.html>> czy D. Turnbull, *Bibliometrics and the World Wide Web*, <<http://www.ischool.utexas.edu/~donturn/research/augmentis.html>>.

**Piotr Nowak**

*WWW as a source of changes in the bibliometric paradigm.  
The beginnings of webometry*

Changes in the bibliometric paradigm triggered by document delivery services via the Internet are presented in the article. Particular emphasis is put to the presentation of the basis for the new method of the quantitative analysis of the Internet, the so-called webometry. The article also presents the author's standing towards a possible reappraisal and reassessment of the current notional apparatus in bibliometry within webometric research. Future applications of bibliometric rules in relation to sources furnished by networked sources are discussed and new areas of theoretical considerations of webometry are outlined.

**Mariusz Nowak**

*Lean management – modern approach to management  
in libraries*

Lean management is a strategy of management worked out in Japan and successfully implemented for the first time in Toyota company. In general, it is such an organization of work that imposes on all employees of a company a duty of constant and particular attention to costs diminishing process and permanent improvement of the quality of offered products. Due to its flexible structure, companies that employ this strategy can, without much shake-up, adjust themselves smoothly to the changeable conditions they have to operate in. In the author's opinion, these rules should also be considered by managers of libraries and their introduction would help better adjust modern libraries to react to changing and varied expectations of their users.

**Artur Jazdon**

*The place of the library in the introduction  
of media technology (medialization) of science*

Medialization of science is understood by the author as the functional and conceptual reinterpretation of the achievements of science and its problems through media and the use of technology. We live in the times characterized by increasingly more important role of media, and, according to experts, this development will go on. We are heading towards a medial society and the fact should not be ignored. However, the image of science that has appeared in the media so far creates many reservations – there is too much chase for cheap sensation, serious problems are presented too superficially and in a shallow way. This is where the library may step in and change the disadvantageous phenomena. Libraries can, for their own benefit and the benefit of their users, become reliable intermediary between the world of science and the world of media and furnish their patrons with information that is verified, credible and reliable. All this can be done with the help of up-to-date technology and the most sophisticated hi-tech information carriers. In strategy projects for the near future, libraries should take the fact into account as the presence of media will become something ever-present in every-day library practice.