

# БИБЛИОМЕТРИЯ И НАУЧНИ ИЗСЛЕДВАНИЯ – НОВИ ПРИЛОЖЕНИЯ НА ТРАДИЦИОННИ МЕТОДИ

*Кристина Върбанова-Денчева*

*Университет по библиотекознание и информационни технологии*

*email: k.dencheva@unibit.bg*

**Резюме:** В доклада се представя приложението на традиционните библиометрични методи в съвременните изследвания и възможностите, които създават новите инструменти за достъп до информационни ресурси в онлайн среда. Накратко са представени учените, които създават най-широко използваните библиометрични подходи, сферата на техните приложения и потенциалът за използване в анализи и оценки на развитието на научните изследвания, иновациите и обвързването им с образованието. Показани са съвременните изследвания на приложението на библиометричните методи в Европа и в света.

**Ключови думи:** библиометрия, библиометрични методи, научни изследвания, качествени оценки

## 1. Въведение

Краткият исторически обзор на развитието на библиометрията очертава няколко етапа от разширяване и обогатяването на нейния обхват в контекста на развитието на научните изследвания. На първо място приложението на математически и статистически методи за оценка на документалния поток е продиктувано от необходимостта от съставянето на оценки за развитието на научните изследвания, които да отразяват възможно най-точно промените в приоритетите и необходимостта от съсредоточаване на осигуряващи условия за тяхното развитие: научен потенциал, научни политики, не на последно място - финансово осигуряване на цялостния изследователски процес.

Обособяването на полето на научните изследвания на библиометрията, а също и на изследователските методи е процес на приложение на математически и статистически методи, а също и на адаптация на методи от други области: икономиката, демографията и др.

## 2. Библиометричните изследвания през XX век – учени, методи и приложения

В съответствие с водещата роля на европейската наука и първоначално осъзнатата необходимост от създаване на единен информационен масив, съдържащ всички библиографски описания на всички публикувани документи във всички области на знанието през 1895 г. в Брюксел Пол Отле и Анри Лафонтен, създават Международния библиографски институт. Този акт е първата стъпка към институционализиране на изследванията на динамиката на развитието на науката и създаването на нови подходи за анализи и оценки.[1]

Библиометрията е въведена като термин от Пол Отле през 1934 г., като в това понятие изследователят е влагал преди всичко статистически методи за наблюдения на документалния поток, състоящ се от систематизираните библиографски описания на новопубликуваните документи и съответно изследвания за определен период от време, тематични области, автори и др.

Едва 33 години по-късно Алън Причърд прилага на практика идеята на Пол Отле, като разграничава статистическите показатели за изследване от библиометричните. Като първо обособено като библиометрично изследване може да се отнесе проведеното през от 1970 г. от Алън Причард и Г. Витиг изследване, което включва съставяне на библиография от повече от 600 заглавия по проблемите на количествените изследвания на документални потоци за периода 1874 – 1959 г. [2]

Обектът на библиометричните изследвания са публикации, групирани по различни признаци: автори; тематични рубрики, езици, страни, години.

Предметът на библиометричните изследвания еволюира в съдържателен аспект от изследване на големи масиви вторична информация до количествени изследвания и разработване на показатели за оценка и класификация на информацията в документални потоци, отразяващи състоянието на науката в цялост и на отделни нейни области.

Целта е да бъдат идентифицирани тенденции, да се осъществява мониторинг на научните изследвания и на средствата и подходите за тяхното оповестяване и поддържане на научна комуникация.

*Хронологично и съдържателно библиометрията преминава в два етапа на развитие: проста библиометрия и структурна библиометрия.*

В периода от 1960 до 1970 г. библиометричните изследвания са съсредоточени в анализирани на библиографски масиви на базата на количествени характеристики за оценка и класифициране по различни категории: по географски признак, хронологичен, по автори, език, възприета класификационна система.

Вторият етап от развитието на библиометрията – след 1970 г., е пряко свързан с интензифициране на научните изследвания, разрастването, появата и развитието на нови научни подобласти, както и оценката на научната ефективност на учени, публикации, издания, институции. Към инструментариума за изследване се добавят показатели за идентификация и оценка на връзки между обектите, тяхната корелация и класификация. Този етап се характеризира с все по-тясното обвързване на библиометрията с наукометрията, чийто обект на изследване е получаване на структурна (качествена) картина и оценка на състоянието на науката.

Картографирането на науката е етап от развитието на науката – поява и развитие на нови информационни материали след 1990 г. – създаването на карти и атласи на науката, разработваните от ISI от 1973 г. бази от данни за цитируемост и коцитируемост на научни публикации, както и разработване на измерители за коцитирането като база за клъстери на масиви от публикации. Приложението на идеята за коцитирането за определянето на силата на връзките между публикациите се превръща в ефективен инструмент за очертаване на нови изследователски полета в научните изследвания и проследяване на тяхното развитие. [8]

## 2.1. Основни направления в структурната библиометрия

Развитието на библиометрията след 1970 г. се очертава в следните основни изследователски полета.

1. Терминологична система на научното знание – създаване на концептуална рамка на корпуса на науката и научния език.

2. Опити за създаване на изкуствен модел на езика на науката.

Дискусиите във философията, по отношение на създаването на пълноценен работещ модел на език, с който може да се оперира в реалната наука, се фокусират в решаването на следните основни проблеми: [2]

а) Да се определи репертоар на лексически единици, използвани в езика на дадена наука;

б) Да се идентифицират системните отношения между лексическите единици.

За създаването на терминологична система на научното знание се използват информационните ресурси, както от библиометрията - вторичните информационни източници, каквито са, например, реферати на публикации, представени в съвременни библиографски бази от данни. В тях информацията е представена в по-компактен вид с филтриране на лексическите единици, които не се съотнасят към основния езиков масив. За тази цел се използват предметни показатели към базите данни, които са специфични за научните изследвания в дадена област и те отразяват лексиката на публикациите в нея.

## 2.2. Изследователски подходи и учени в областта на библиометрията през XX век

Очертаването на полето на изследвания на библиометрията е тясно свързано и с учени научни школи, които създават най-широко използваните библиометрични подходи, сферата на техните приложения и потенциалът за използване в анализи и оценки на развитието на научните изследвания, иновациите и обвързването им с образованието.

Един от основоположниците на библиометрията е **Самюел Брадфорд**, който през 1934 г. формулира Закон за разсейване на информацията. [3]

Изследванията на С. Брадфорд имат за резултат установяването на следните зависимости:

- Между броя на авторите, публикуващи по дадена тема и броя на изданията (списанията) в тази област;
- Между броя на публикациите и броя на техните автори.

Според Закона за разсейването на информацията съществуват и следните закономерности при индексиранието на информацията:

- Между обема на документите в даден информационен масив и броя на необходимите речници, които се използват за тяхното индексирание;
- Логаритмична зависимост между броя на необходимите речници за индексиранието на даден обем документи в информационния масив.

Други установени зависимости са между разпределението на статиите в периодичните издания, които се разпределят в 3 зони: статии в издания в дадената тематична област (т.нар. „ядро“) –  $r_1$ ; статии в издания с по-широк тематичен профил –  $r_2$ ; статии в популярни издания –  $r_3$ .

Съотношенията между броя на изданията се подчинява на следната закономерност:

$r_1:r_2:r_3 = 1 : a : a^2$ , където  $a$  е съотношението:  $r_3:r_2 = r_2:r_1 = a$

Приложение на закона на С. Брадфорд за разпределение на броя на заглавията по дадена тема (J) и броя на темите, разпределени в групи – (K)

J – брой на заглавия по дадена тема, който се разделя на k групи, всяка от които съдържа  $j_1, j_2, j_3, \dots, j_k$  заглавия при което всяка група съдържа еднакъв брой статии по дадена тема, тогава е в сила зависимостта:

$$j_i = b k^{j_i-1} = b k^{i-1} j_1, i = 1, 2, 3, \dots, m;$$

$j_1$  – брой на списанията в ядрото,  $b k > 1$  – множител на Брадфорд за разделянето на j заглавия на k групи. Множителят bk намалява по стойност, когато се увеличи броят на зоните – k.

Друг изследовател, който установява зависимости между честотата на срещане на думите в даден текст и идентифициране на тематичната област на текста е законът (правилото) на **Джордж Ципф**. Законът е известен още като „Принцип на най-малкото усилие и/или пътят на най-малкото съпротивление“ и намира приложение в конкретни изследвания в лингвистиката (например - лексикална дифузия, учене на език, анализ на разговор). При измерването на честотата на използване на дадена дума в даден език и по-специално изследването на разпределението на отделни думи в големи откъси от текстове се получава интересна зависимост между думите и честотата на тяхното присъствие в текстовете. При подреждането им в низходящ ред на честотата на тяхното появяване, присвоения пореден номер в тази последователност остава постоянна величина.[4]

Честотното разпределение на думите по Ципф включва последователното подреждане на думите от голям откъс от текст, следвайки намаляването на честотата на тяхното появяване в текста, ако се присвои число като пореден номер на честотата на появяването в нарастващ ред, това число остава постоянна величина за всяка дума и се нарича *ранг*.

Това правило Ципф апробира не само в лингвистиката, но също и в психологията, за която го модифицира в следното правило: „Във всяко човешко действие се влага най-малко усилие, за да се изпълни дадена задача“. Приложение правилото на Ципф намира също и в социологията, икономиката, маркетинга и в информационните науки.

Законът на Ципф е следствие от общия „принцип на най-малкото съпротивление“ – най-често срещаните думи на всеки език са кратките служебни думи. Тяхната употреба изисква най-малки усилия, което обяснява голямата честота на появата им. Законът на Ципф може да се прилага с известна адаптация, коригираща не точно най-малкото съпротивление.

Математическото описание на закона на Д. Ципф се представя със следната зависимост

$C = F(i) * r(i)$ , където c е константа за всяка дума, която се среща с определена честота F, а r е присвоеният ранг в зависимост от честотата.

За универсално приложимост на зависимостта по закона на Д. Ципф за различни типове данни, е установено точно съответствие, което може да се представи с промененото уравнение:

$$f(i) = c(r(i)-a)^{-s}, \text{ където } a, c \text{ и } s \text{ са константи, съответно:}$$

a е константа, която коригира съпадението за общоупотребяваните думи, които имат нисък ранг;

c е константа за думи с по-висок ранг;

s – константа за думи с висок ранг – по-рядко срещани думи.

Законът на Д. Ципф се прилага в библиотекознанието за автоматизирано класифициране на библиотечни документи, в библиометрията и в наукометрията.

От изследователите, които провеждат систематични проучвания на документалните потоци е **Алфред Лотка**. Той установява зависимост между разпределението на авторите и броя на публикуваните им статии, която е известна като „законът на обратния квадрат“. А. Лотка изследва честотата на публикациите на автори в дадена тематична област и установява, че броят на авторите, които имат n на брой публикации е равен на  $1/n^2$  от броя на авторите, които имат само една публикация. В идеалния случай това означава, че 100% от публикациите се формират от: 60% автори на една публикация; 15% автори на 2 публикации; 7% автори на 3 публикации и 6% автори на повече от 10 публикации. [5]

Законът на Лотка се изразява с формулата: [6]

$$f(x) = k / x^a, \text{ където } k \text{ и } a \text{ са константи, } > 0; x \text{ е броят на авторите с } x \text{ публикации (} x = 1, 2, 3, 4)$$

А. Лотка установява, че за съвкупност от 1000 автори константите имат стойности:  $a = 2, k = 0.608$ .

На тази база се прави прогноза колко автори (x) ще имат (x) публикации – колко автори ще имат 1 публикация, 2, 3 и т.н.

Изследванията на А. Лотка са в широкия спектър на създаването на математически модели за развитие на биологични системи, включително в екологията и за проследяване на динамиката на населението. Принципът на Лотка, приложим и почти един век след формулирането му, гласи: „...оцеляват и просперират тези организми, които улавят и използват енергията по-ефективно от своите конкуренти.“ [4]

*Идеите на А. Лотка са доразвити от В. Волтера и зависимостите са известни като „уравненията на Лотка-Волтера“. [7]*

Друг изследовател, който установява зависимости в документопотоците и има принос в развитието на библиометрията е **Вилфредо Парето**. Той формулира през 1906 г. Закон (Правило) на Парето (80:20), а през 1926 г. съвместно с Г. Моска публикуват фундаменталния за политологията и за социологията труд „Теория на елитите“. [9]

Закон на Парето се прилага в библиометрията за установяване на зависимостта между броят на публикациите и техните автори. Според принципа на В. Парето на 20% от авторите принадлежат 80% от публикациите, съответно и броят на учените и техните открития се подчинява на същата зависимост.

Съотношението на Парето 20: 80 може да бъде в диапазона от 5: 95 до 30: 70 в зависимост от различни и много на брой фактори, специално за броя на изследователите и броя на техните публикации съществен фактор е продължителността на творческия им път.

Законът на В. Парето намира широко практическо приложение в социологията, мениджмънта и маркетинга, в прогнозирането в най-разнообразни сфери – от бизнеса до инженерните науки и научните изследвания.

Едни от общовалидните следствия от закона на В. Парето са:

- Приблизително 80% от богатството принадлежи на 20% от населението;
- Приблизително 80% от работата се извършва от 20% от работниците;
- Приблизително 80% от стоката се купува от 20% от купувачите;

В проведените изследвания от В. Парето в широк диапазон от приложни дейности са установени следните зависимости:

- Повечето от неприятностите възникват поради действието на малък брой силно разрушителни сили;
- Преобладаващият брой дефекти и свързаните с тях загуби произтичат от относително малък брой причини;
- Повечето действия (групови или индивидуални) са загуба на време, тъй като нямат значение за постигане на желания резултат.

### **3. Изводи/Дискусия**

Приложението на Закона на С. Брадфорд за разсейването на информацията има съвременно приложение за установяване на „ядра“ от издания, които формират и най-авторитетните (с висок импакт-фактор) издания.

Съчетаването на закона на С. Брадфорд с установената зависимост от Д. Ципф позволява да се състави точен математически модел за нарастването на публикациите в дадена тематична област и съответно прогноза за развитието в бъдеще за възникване на нови изследователски полета.

Приложението на Закона на Д. Ципф за съставяне на терминологичния речник на науката създава възможност за точното отчитане на честотата на всички термини. То подпомага създаването на математически модели, които идентифицират неслучайния характер на съвместното срещане на термините.

Решаването на тези задачи води до следните резултати:

- а) проследяване на количествената динамика на групи термини в научната област;
- б) очертаване на специфичните връзки между термините – създаване на тезаурус на науката – статични и динамични.

Развитието на изследванията за създаване на терминологична система на науката продължава с въвеждането на подходи за семантична и социологическа интерпретация на получените резултати, което е във възможностите на изкуствения интелект. Това позволява очертаването на динамиката на разпределението на публикациите по определени рубрики и идентифицирането на приоритетни направления в научните изследвания в краткосрочен и дългосрочен план.

Приложението на закона на А. Лотка в съчетание със законите на С. Брадфорд и Д. Ципф също подпомага прецизирането на базата за разработване на прогнозен модел за развитие на публикационната активност и съответно за нейното стимулиране в различни аспекти: тематична област, възраст на изследователите, разпределение по пол, език, региони и др.

Основен проблем на библиометрията на ХХ век е изследването на информационните потоци в тяхната статика – материализираните документални потоци отразяват динамиката на научните изследвания и съответно проследяването им следва модела на епохата до интернет и технологиите на офф лайн комуникацията. Като обект на изследване от библиометрията са статистическите разпределения в процесите, свързани с дисперсията, разсейването и използването на единиците информация. Невъзможността да се реши основния проблем на библиометрията със средствата и подходите до 70-те години на ХХ век - да се създаде теоретична база с формулиране на закони от общ характер за описание на най-различни структурни форми и функционални аспекти на обработката на информацията, се преодолява през ХХI век с технологиите на изкуствения интелект, в чийто създаване се използват и постиженията на

библиометрията. Не на последно място трябва да се отбележи и напредъка в наукометричните изследвания и тяхната интеграция с библиометрията, придвижващи библиометричните изследвания на следващия – структурен етап от тяхното развитие. [8]

Последващите статистически анализи на резултатите от изследванията са база за конкретизиране на изводите за значимостта на изследвания обект, продуктивността на учените, научната ефективност на научните публикации (цитируемост), научния потенциал на дадена страна, провежданата политика за научни изследвания.

#### 4. Заключение

Значението на представените закони е, че техните основатели поставят основата за систематично проучване на документопотоците и съответно изиграват значителна роля в развитието на библиометрията. Използването на тези методи за управление на масиви от списания или за оценка на научното влияние на периодични издания, автори, статии, изследвания е изиграло своята роля за получаване на обективна информация за ефективността на научната дейност и информационното подпомагане на науката през миналия век.

Практическо приложение на библиометричните подходи в съвременността е резултат от тяхното адаптиране към динамиката на промените в информационните носители, структурата на информационните потоци, подходите за тяхното изследване и най-вече възможностите, които предоставят съвременните технологии.

Изследването на информационните потоци и установяването на закономерности, включително и с използване на представените изследователи и техните закони, дават възможност да се подпомогне мениджмънта на библиотеките, включително дигитални, създаване на дигитални архиви и хранилища, прогнозиране на развитието на учени, научни школи и институции, не на последно място – обективизиране на критериите за предприемане на мерки за стимулиране на научните изследвания, провеждане на целенасочена научна политика за страни и региони. В условията на интензивни изследвания и постигнати резултати в разработването на изкуствен интелект и множеството интелигентни приложения, са дали своя принос и С. Брадфорд, Д. Ципф, А. Лотка, В. Парето, чийто постижения в установяването на зависимостите в създаването и развитието на информационните потоци имат голямо прогностично значение за създаване на модел за развитие на информационната среда и в бъдеще. В това отношение все по-силна става синергията между библиометрията и наукометрията, които използват еднаква законова и теоретична база, доразвита и подпомогната от развитието на компютърните науки.

В обхвата на приложението на резултатите от изследванията са както възникнали социални потребности, така и глобални проблеми за човечеството като търсене и навлизане в енергетиката на нови енергийни източници, пробив в лечението на тежки заболявания, продължение на изследванията в полето на нови научни открития и тяхното приложение в иновации, не на последно място – потенциалът за налагане на нови тенденции в развитието на хуманитарните науки.

#### Литература

1. Младенова, Мария. Библиотечни класификации. Изд. НБ „Иван Вазов“, Пловдив, 1996, с. 111 – 115.
2. Маринов, Марин. Логиката на науката и методологията на Карл Попър. УИ „Св. Кл. Охридски“, 1992., с. 211- 219.
3. LAWANI, S. M.. "Bibliometrics: Its Theoretical Foundations, Methods and Applications" Libri, vol. 31, no. 4, Jahresband, 1981, pp. 294-315. <https://doi.org/10.1515/libr.1981.31.1.294> (8.05.2022, 12.00)
4. Zipf, G.K. Human behavior and the principle of least effort: an introduction to human ecology (1902 1950) / G.K. Zipf. - Cambridge: Addison-Wesley Press, 1949. -573 p.
5. Kingsland, Sh. Alfred J. Lotka and the origins of theoretical population ecology. // <https://www.pnas.org>, 10.10.2021, URL: <<https://www.pnas.org/content/112/31/9493>> <https://www.pnas.org/doi/10.1073/pnas.1512317112>> (5.05.2-22, 21.00)
6. Lotka, A.J. The Frequency Distribution of Scientific Productivity. / A.J. Lotka // Journal of the Washington Academy of Science. - Vol. 16 (12). - P. 317-323.
7. Strauss, R. Augmenting Neural Differential Equations to Model Unknown Dynamical Systems with Incomplete State Information. <<https://arxiv.org/abs/2008.08226>> (6.06.2022, 20.00)
8. Brookes, B.C. Towards informetrics: Haitun, Laplace, Zipf, Bradford and the Alvey programme / B.C. Brookes // J. Doc. - 1984. - Vol. 40. - № 2. - P. 120 – 143.

9. Георги Фотев. Социологическите теории на Е. Дюркем, В. Парето, М. Вебер. София: Наука и изкуство, 1979, с. 258 - 276с.

## **BIBLIOMETRIC AND SCIENTIFIC RESEARCH - NEW APPLICATIONS OF TRADITIONAL METHODS**

***Professor Kristina Varbanova-Dencheva, PhD***

**University of Library Studies and IT, k.dencheva@unibit.bg**

**Abstract:** *The report presents the application of traditional bibliometric methods in modern research and the opportunities created by the new tools for accessing information resources in an online environment.*

*The scientists who created the most widely used bibliometric approaches, the scope of their applications and the potential for use in analyzes and evaluations of the development of scientific research, innovation and their connection with education are briefly presented. Contemporary studies of the application of bibliometric methods in Europe and the world are shown.*

**Keywords:** *bibliometrics, bibliometric methods, scientific research, qualitative assessments*