

УДК 004.942

МОДЕЛИРОВАНИЕ ПРОЦЕССА ОБРАЩЕНИЯ К ЭЛЕКТРОННЫМ ИНФОРМАЦИОННЫМ ИСТОЧНИКАМ НА ОСНОВЕ ИНФОРМЕТРИЧЕСКОГО ЗАКОНА БРЭДФОРДА

В.В. Писляков

Аннотация

Рассматриваются вопросы использования информетрического закона Брэдфорда при моделировании обращений к электронным журналам. Два варианта модели Брэдфорда применяются к массиву экспериментальных статистических данных. Строится график рассеяния спроса по источникам, определяется компактное «ядро» наиболее востребованных изданий.

Введение

Методы информетрии, описывающие закономерности производства и потребления информации, активно применяются при моделировании и анализе различных аспектов научной деятельности [1]. Основополагающие информетрические законы Брэдфорда, Парето, Лотки, Ципфа и др. находят подтверждение, будучи примененными к самым разнообразным информационным процессам: рассеянию опубликованных статей по журналам и авторам, цитированию научных публикаций, чтению монографий и периодических изданий.

Закон Брэдфорда имеет две формулировки: «вербальную» и «графическую». Согласно первой, если выделить множество журналов, в которых появляются статьи, посвященные некоторой выбранной научной тематике, и упорядочить их в порядке убывания количества данных статей – от изданий, в которых имеется наибольшее количество публикаций исследуемой тематики, до тех, в которых встретилась лишь одна статья, относящаяся к выбранной дисциплине, – то это упорядоченное множество журналов можно разбить на три зоны так, чтобы в каждой было одинаковое суммарное количество статей по заданной теме, при этом количество журналов в каждой зоне будет относиться друг к другу как $1 : q : q^2$, где q – некоторое число, большее единицы, которое будем называть коэффициентом Брэдфорда [2, 3]. В дальнейшем закон был распространен на произвольные n «зон Брэдфорда», которые содержат равное число статей, а связь между количеством журналов в них, соответственно, задается пропорцией $1 : q : q^2 : \dots : q^{n-1}$.

Графическая формулировка закона Брэдфорда, также содержащаяся в [2], проиллюстрирована на рис. 1. Вновь упорядочим множество журналов в порядке убывания числа релевантных статей (в порядке убывания их «продуктивности» по заданной тематике). Отложим по оси абсцисс натуральный логарифм числа первых s наиболее продуктивных журналов, а по оси ординат – кумулятивное (суммарное) число релевантных статей i , содержащихся в этих s изданиях. Тогда, согласно графической формулировке закона Брэдфорда получившаяся кривая в полулогарифмических координатах (ее называют «библиографом») будет вогнутой

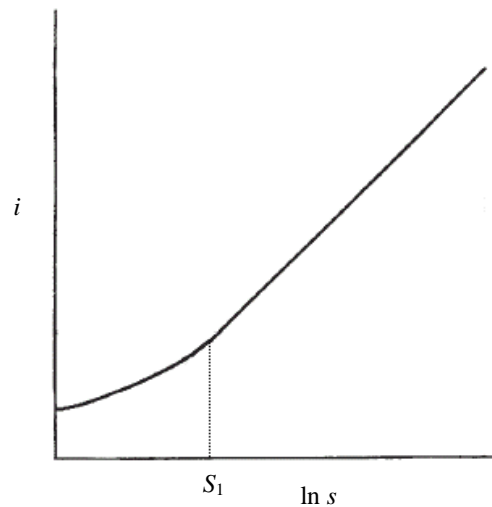


Рис. 1. Библиограф: график зависимости кумулятивного числа статей в s первых журналах от $\ln s$

до некоторого значения $s = S_1$, после которого она перейдет в прямую (рис. 1). Именно S_1 предлагается брать в качестве первой зоны Брэдфорда.

Конкретные формулы для двух фрагментов библиографа предлагались, например, в [4]. Кроме того, во многих исследованиях отмечался переход верхней части прямолинейного отрезка в выпуклую кривую, то есть отклонение библиографа от прямой вниз. Данный эффект получил название Groos Droop [5], и ему обычно предлагались объяснения, связанные с неполнотой рассматриваемой библиографии, то есть неучтенными журналами, публикующими по несколько (1–2) релевантных статей (см. [4], возражение на это содержится в [6]), или с эффектами соединения нескольких узкоспециализированных библиографий, в точности подчиняющихся закону Брэдфорда, в единый политематический массив [7].

Закон Брэдфорда в различных его формулировках применялся к распределению по изданиям научных статей широкой тематики: по геофизике и смазочным материалам [2], по мышечным волокнам и компьютерным наукам [4], по исследованию операций [8], по термофизическим свойствам веществ [9] и др. Кроме того, было высказано предположение, что закон применим к информационным процессам иной природы, и с различной степенью успеха его проверяли на распределении статей по авторам [10] и странам публикаций [11], на распределении *сделанных* цитирований по журналам [12, 13] или *полученных* цитирований по авторам [14, 15].

К теме настоящей работы непосредственно примыкают исследования применимости закона Брэдфорда к моделированию процессов *потребления* научной информации – изучение статистики обращений к информационным источникам различной природы. Данный аспект исследовался в приложении к книговыдачам [16] и выдачам периодики [17] в библиотеках, к использованию реферативных периодических изданий [4], к массиву заказов на периодику, поступивших в службу доставки документов [18].

Несмотря на достаточное количество трудов, применяющих закон Брэдфорда к различным ситуациям и информационным процессам, нам неизвестны исследования, посвященные приложению данного информетрического распределения к моделированию использования *онлайн-баз данных научной периодики* (библиотек электронных журналов). Это представляется серьезным пробелом, кото-

рый настоящая работа ставит себе целью восполнить в результате применения закона Брэдфорда для анализа обращений к зарубежной электронной периодике в университетской библиотеке (библиотека Государственного университета – Высшей школы экономики, ГУ-ВШЭ), а конкретно – для анализа рассеяния использования по наименованиям источников. Отметим, что ранее автором на аналогичном статистическом массиве для исследований использовался другой информетрический закон, принцип Парето, применимость которого была доказана в [19].

Еще одним приложением закона Брэдфорда во многих упомянутых работах, начиная с исходной [2], является определение «ядра» литературы, концентрирующего в себе журналы, на которые приходится основная часть релевантных статей. При использовании вербальной формулировки закона «ядро» считается совпадающим с первой зоной Брэдфорда, определенной математически. При более распространенном для данной задачи графическом подходе ядро ограничивается точкой перехода вогнутого участка библиографа в прямую. В случае распределения опубликованных статей по журналам таким образом определялся состав ядра периодических изданий по соответствующей дисциплине. В случае изучения процесса чтения книг или заказов на доставку документов выявлялись «ядерные» издания, представляющие ключевой интерес для читательской аудитории. Мы ставим также себе целью, применив информетрическую модель Брэдфорда, определить ядро читательского спроса аудитории ГУ-ВШЭ на научную информацию, установить множество наиболее востребованных источников.

1. Условия исследования

Электронный фонд библиотеки ГУ-ВШЭ на момент сбора статистических данных (2004 г.) состоял из изданий, размещенных на более, чем десяти различных онлайн-платформах. Наиболее активно использовались ресурсы периодики EBSCO, JSTOR, ProQuest и ScienceDirect, которые и были взяты для настоящего исследования. Хотя все четыре системы предоставляют доступ к электронным изданиям, они отличаются по своим задачам, структуре контента, объему доступной информации, равно как по другим параметрам, что обеспечивает широкий и разносторонний охват электронных источников. Данные по составу и объему баз данных, которые были доступны для пользователей ГУ-ВШЭ в указанных ресурсах, собраны в табл. 1. Рассматривается полнотекстовый материал, так как именно его использование будет учитываться в настоящем исследовании (см. далее). В общей сложности, за вычетом дубликатов (наименований, представленных в двух или более базах одновременно), четыре ресурса охватывают около 13400 источников в полных текстах.

2. Методика

Каждая из четырех рассматриваемых электронных библиотек собирает и предоставляет статистику активности пользователей каждой организации с распределением спроса по наименованиям периодических изданий. Эта статистика считается в автоматическом режиме и свободна от серьезных технических и методологических сложностей подсчета обращений к традиционным печатным изданиям (особенно на уровне отдельных статей).

В ресурсах содержатся не только полнотекстовые журналы, но также и издания, доступ к которым предоставлен лишь на уровне библиографических сведений о статьях или их аннотаций. Тем не менее, в настоящей работе нас будут интересовать только обращения к полным текстам как наиболее полноценная и поддающаяся интерпретации статистика использования.

Табл. 1

Полнотекстовые ресурсы, участвующие в настоящем исследовании

Платформа	Доступные полнотекстовые базы данных	Всего наименований источников
EBSCO	Academic Search Premier, Business Source Premier, MasterFILE Premier, Newspaper Source, Regional Business News, Health Source	около 11000
JSTOR	Collections: Arts & Sciences I, Arts & Sciences II, Business, Language & Literature	416
ProQuest	ABI/INFORM Global, Academic Research Library, ProQuest Education Journals, Social Sciences PlusText, ProQuest Psychology Journals, ProQuest Computing	около 3600
ScienceDirect	Business, Management and Accounting + Backfiles; Economics, Econometrics and Finance + Backfiles; Social Sciences + Backfiles	310

Для единообразного представления результатов анализа возьмем в роли основного показателя использования ресурсов число открытых полнотекстовых статей, подсчитываемое всеми четырьмя ресурсами и рекомендуемое международными экспертами в качестве одной из главных метрик при оценке востребованности документарных баз данных. При этом многократное обращение к одной и той же электронной публикации подсчитывается каждый раз – по числу отдельных обращений (в отличие от подсчета загрузок «уникальных» полнотекстовых документов).

Данные по использованию публикаций объединены по всем четырем ресурсам, с суммированием показателей по наименованиям, представленным более чем в одном из них. Журналы идентифицировались по ISSN, при отсутствии ISSN идентификация проводилась по наименованию. Издания с различными ISSN и одинаковым заглавием считались разными. Не проводилось объединение данных по одному и тому же журналу, если он в течение своей истории менял ISSN или наименование.

В качестве временного промежутка, на котором собиралась статистика обращений к базам данных, выбран 2004 г., с января по декабрь включительно. За данный период было зафиксировано 41959 обращений к полным текстам статей из 2590 уникальных наименований. Такой масштаб исследования обеспечивает надежность и статистическую значимость его результатов.

3. Результаты

Доля изданий, к которым было хотя бы одно обращение, составила от 10% для EBSCO до 76% для ScienceDirect, в целом по всем ресурсам было востребовано 2590 из около 13400 журналов, то есть 19%. Наблюдалась высокая концентрация спроса на небольшом числе изданий (подробнее см. [19]), как исходно и предполагает закон Брэдфорда. Наибольшее число обращений было к журналу *American Economic Review*: 1306 открытых статей (3.1% от всех загрузок), это число обозначим через s_1 , оно потребует в дальнейших выкладках.

Сначала обратимся к графической формулировке закона Брэдфорда и построим библиограф для имеющейся статистики (интегральной по всем четырем ресурсам), отразив зависимость кумулятивного числа востребованных статей от логарифма количества журналов, которые упорядочены в порядке убывания спроса (рис. 2).

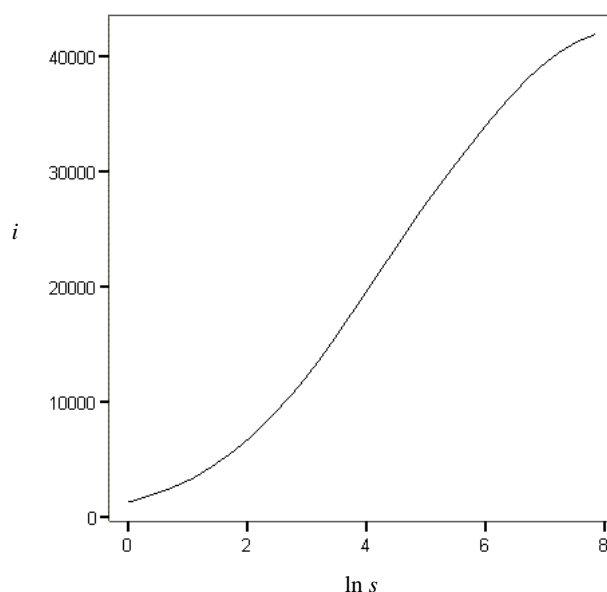


Рис. 2. Библиограф для распределения обращений по наименованиям журналов

Полученная кривая имеет классический вид: выделяется начальный фрагмент, имеющий положительную вторую производную, далее участок, близкий к прямолинейному, и в конце – отклонение от прямой в виде выпуклого участка библиографа, Groos Doop. Таким образом, в рамках настоящего исследования на основании имеющегося статистического массива можно говорить о хорошем выполнении графической формулировки информетрической модели Брэдфорда для процесса обращения к электронным информационным источникам.

Обратимся теперь к исследованию применимости модели Брэдфорда с использованием ее «вербальной», математической, формулировки. Для этого нам необходимо выделить зоны Брэдфорда и изучить соотношения между количествами журналов, попадающих в них.

Прежде всего воспользуемся элементарным методом, который предполагается классической формулировкой закона: разобьем всё множество обращений к статьям на три равные части и определим, какое число журналов попадает в каждую из частей. Мы имеем 41959 обращений, то есть по 13986–13987 в каждой из трех зон. Исходя из этого и опираясь на экспериментальные данные распределения спроса по наименованиям журналов, нами получено, что первую зону Брэдфорда формируют 26 журналов (из каждого открыто от 1306 до 275 публикаций, всего – 13987 обращений), во вторую зону входят 139 изданий (от 272 до 43 обращений к каждому, всего – 13975), в третью – 2425 журналов (от 42-х до 1-го запроса на издание, всего – 13997). Результаты представлены в табл. 2.

Из анализа табл. 2 можно заключить, что построенная «классическим» методом информетрическая модель Брэдфорда плохо соответствует рассматриваемым эмпирическим данным: числа журналов в трех зонах не образуют геометрической прогрессии, и коэффициент Брэдфорда q (отношение числа журналов в n -й зоне к числу в $(n-1)$ -й), который при идеальном выполнении закона должен оставаться постоянным, варьируется более чем в три раза, от 5.4 до 17.5.

Воспользуемся более развернутым методом определения параметров информетрической модели Брэдфорда (то есть количества зон n , коэффициента Брэдфорда

Табл. 2

Разбиение на три зоны Брэдфорда методом простого деления множества статей на равные части

Зона Брэдфорда	Число журналов в зоне	Процент журналов	q	Число статей в зоне
I	26	1.0%	–	13987
II	139	5.4%	5.35	13975
III	2425	93.6%	17.45	13997
всего	2590	100%	–	41959

q и др.), предложенным Л. Эгге, который не делит зонами все статьи на одинаковые доли, наблюдая затем за флуктуациями q , а, наоборот, фиксирует q и n , стремясь найти их такими, чтобы заданное ими разбиение на зоны Брэдфорда давало приблизительно равное число статей в каждой зоне, пытаясь также при этом устранить погрешности, возникающие из-за того, что число журналов в каждой зоне должно быть целым. Вывод и основания этой методики изложены в [20], мы проиллюстрируем работу метода для $n = 3$, а затем сразу приведем результаты для других значений n .

Определим методом Эгге коэффициент Брэдфорда для $n = 3$:

$$q = (e^\gamma s_1)^{1/n} \approx (1.781 \cdot 1306)^{1/3} \approx 13.25$$

где γ – постоянная Эйлера–Маскерони, s_1 – число обращений к самому спрашиваемому журналу. Далее вычислим величину S_1 первой зоны Брэдфорда:

$$S_1 = S \frac{q-1}{q^n-1} = 2590 \cdot \frac{13.25-1}{13.25^3-1} \approx 13.65$$

где S – суммарное число журналов, к которым были обращения. Поскольку значение S_1 получается не целым, используем округление до ближайшего целого $\{S_1\} = 14$ – размер первой зоны (здесь нами введена небольшая поправка к методу Л. Эгге [20], который предлагает округлять строго в сторону меньшего целого, однако наши эксперименты показали худшую эффективность такого подхода). На первые 14 наиболее востребованных журналов приходится 10049 открытых статей.

В двух остальных зонах число журналов вычисляется как результат округления до ближайшего целого произведений qS_1 и q^2S_1 , что дает 181 и 2396 изданий соответственно; отметим, что вычисленные таким образом зоны Брэдфорда дают в сумме 2591 журнал, что превышает общее число журналов S на единицу. Это произошло в результате накопления ошибок округления, поэтому размер последней группы скорректируем на единицу, приравняв его 2395. По данным статистики обращений на следующие после первой зоны Брэдфорда 181 и 2395 журналов приходится, соответственно, 19063 и 12847 статей. В результате получаем данные, представленные в табл. 3.

Сравним приведенные в табл. 2 и табл. 3 результаты выделения трех зон Брэдфорда двумя различными методами – методом С. Брэдфорда и методом Л. Эгге. Сама техника этих методов ведет к тому, что в первом случае практически одинаково число статей в каждой зоне, но при этом сильно варьируется коэффициент q , во втором – наоборот, отношение q числа журналов в группах друг к другу практически постоянно, однако число статей по зонам значительно (почти в два

Табл. 3

Три зоны Брэдфорда, определенные методом Л. Эгге

Зона Брэдфорда	Число журналов в зоне	Процент журналов	q	Число статей в зоне
I	14	0.5%	–	10049
II	181	7.0%	12.9	19063
III	2395	92.5%	13.2	12847
всего	2590	100%	–	41959

раза) отличается друг от друга. Следовательно, «вербальная» формулировка закона Брэдфорда для имеющихся экспериментальных данных и $n = 3$ выполняется малоудовлетворительно, вне зависимости от того, какой способ разбиения на зоны применяется. Несмотря на сильную концентрацию и неравномерность распределения по журналам обращений пользователей, что предсказывается законом Брэдфорда, математическая форма этой концентрации отличается от описываемой вербальным вариантом закона.

Л. Эгге допускает (и даже рекомендует) разбиение журналов на большее, чем $n = 3$, число зон. Результаты таких разбиений для некоторых n приведены в табл. 4.

Отметим, что при возрастании n уменьшается определяемая по методу Эгге величина S_1 первой зоны Брэдфорда, и, в связи с тем, что в ней должен быть хотя бы один журнал ($S_1 \geq 1$), действует ограничение на число зон n :

$$n \leq \frac{\ln(e^\gamma s_1)}{\ln\left(1 + \frac{e^\gamma s_1 - 1}{S}\right)} \approx 12.4.$$

Таким образом, нет смысла пытаться разбить количество журналов методом Эгге больше, чем на 12 зон.

Анализ последнего столбца (число статей в зонах Брэдфорда) должен показать, насколько хорошо выполняется для исследуемого массива данных вербальная формулировка закона Брэдфорда: в случае идеального его выполнения количество статей в зонах было бы постоянным. Это далеко не так: например, отношение максимального числа статей к минимальному варьируется от 1.90 ($n = 3$, см. табл. 3) до 4.69 ($n = 10$). Интересна динамика количества статей при переходе от первых зон к последним. Сначала их число в каждой последующей зоне увеличивается, затем достигает своего максимума, вокруг которого (в случае больших n) имеются 3–4 зоны с близким друг другу числом статей, после чего в последних зонах количество статей снова заметно уменьшается.

Область, где размер зон (выраженный в статьях) близок друг к другу, – это область приемлемого выполнения вербальной формулировки закона Брэдфорда. Нетрудно заметить, что она совпадает с линейным участком библиографа. Например, при $n = 12$ три зоны VI – VIII содержат практически одинаковое число статей: отклонение составляет менее 7%. В этих зонах содержатся журналы с номерами s от 28 до 195. На библиографе (рис. 2) этому соответствуют точки $\ln 28$ и $\ln 195$, то есть отрезок оси абсцисс [3.3, 5.3], который расположен в той части, где график представляет собой практически прямую линию. При более детальном анализе можно показать, что линейность библиографа – достаточное (но не необходимое) условие для выполнения на соответствующем участке вербальной формулировки закона Брэдфорда. Пока же мы вынуждены констатировать,

Табл. 4

Разбиение на зоны Брэдфорда методом Эгге: варьирование числа зон

Зона Брэдфорда	Число журналов	q	Число статей
$n = 4; q = 6.94; S_1 = 6.62$			
I	7	–	6469
II	46	7.67	12865
III	319	6.93	14088
IV	2218	6.95	8537
$n = 5; q = 4.71; S_1 = 4.14$			
I	4	–	4336
II	19	4.75	8806
III	92	4.84	12165
IV	433	4.71	10516
V	2042	4.72	6136
...			
$n = 10; q = 2.17; S_1 = 1.30$			
I	1	–	1306
II	3	3.00	3030
III	6	2.00	3828
IV	13	2.17	4978
V	29	2.23	6042
VI	63	2.17	6123
VII	137	2.17	5543
VIII	297	2.17	4984
IX	644	2.17	3906
X	1397	2.17	2219
...			
$n = 12; q = 1.91; S_1 = 1.01$			
I	1	–	1306
II	2	2.00	2105
III	4	2.00	3058
IV	7	1.75	3580
V	13	1.86	4210
VI	26	2.00	5075
VII	49	1.88	5055
VIII	93	1.90	4723
IX	176	1.89	4292
X	339	1.93	3874
XI	646	1.91	2951
XII	1234	1.91	1730

что вербальный вариант информетрической модели Брэдфорда плохо описывает исследуемый нами процесс обращений к онлайн-информационным источникам, за исключением отдельного участка в средних зонах. Отметим также, что это напрямую связано с небольшой величиной линейного фрагмента библиографа, достаточно быстро переходящего в Groos Dгоор.

Наконец, вернемся к графической формулировке модели Брэдфорда и определим с ее помощью величину «ядра» литературы, граница которого, напомним, совпадает с переходом библиографа от вогнутой части к прямолинейной.

Хотя в настоящее время нет общепризнанного подхода к определению точки этого перехода (например, про приближение библиографа прямой авторы [21] пишут: «В большинстве литературы данные просто отображаются на графике, и при-

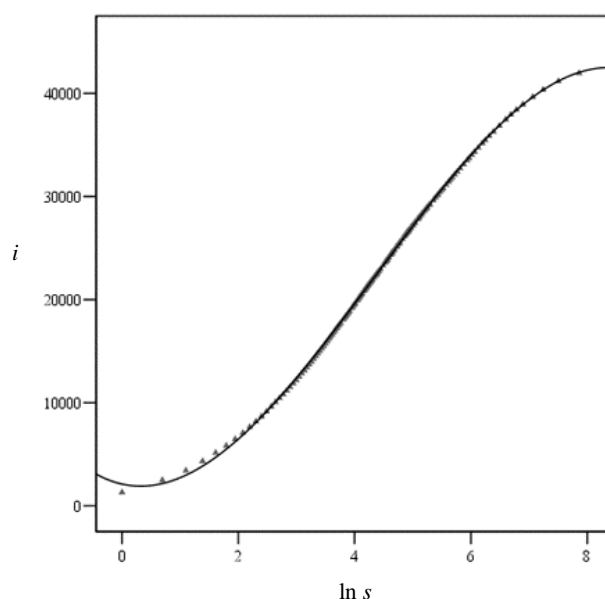


Рис. 3. Приближение библиографа полиномом третьей степени (точки – исходные данные, кривая – кубический многочлен)

ближение осуществляется на глаз. Мы не обнаружили такого подхода к приближению подобных данных, который был бы привлекателен с теоретической точки зрения», мы предложим два метода определения данной границы. Прежде всего, судя по виду графика на рис. 2, можно предположить, что его можно достаточно точно приблизить кубическим многочленом. Такое приближение методом наименьших квадратов показано на рис. 3 (выполнено при помощи программы SPSS 12.0).

Из рис. 3 видно, что приближение практически идеальное, это подтверждает коэффициент детерминации (R^2): он равен единице с точностью до третьего знака после запятой. Теперь найдем точку перегиба аппроксимирующей функции – вблизи нее график имеет квазипрямолинейный вид. Используя полученные коэффициенты аппроксимации ($a_3 = -155.7$ при третьей степени полинома и $a_2 = 2030.1$ – при второй), найдем точку перегиба:

$$x = -a_2/(3a_3) \approx 4.35 \Rightarrow S_1 \approx 77. \quad (1)$$

«Ядро», полученное данным образом, содержит 77 журналов. Естественно, как следует из самого метода, таким способом мы находим «оценку сверху»: дальше этой точки аппроксимирующий полином из вогнутого становится выпуклым, значит, она *уже* находится на прямолинейном участке (и скорее всего где-то в его середине). Итак, ядро спроса литературы, полученное методом Брэдфорда, точно не превышает 77 наименований журналов.

Уточнить искомую границу нам поможет обращение к табл. 4 и вербальной формулировке информетрической модели Брэдфорда. Как мы видели, при $n = 12$ существуют зоны с шестой по восьмую, где закон Брэдфорда практически точно выполняется – количество статей в каждой из этих зон близко к константе. Если предположить, что это сопровождается линейностью библиографа на данном участке (что не является необходимым условием, но в данном случае такое предположение вполне закономерно), то можно сказать, что граница «ядра» приблизительно совпадает с началом четвертой зоны при разбиении на 12 зон. Из табл. 4

следует, что это – точка $x = \ln 28$, и ядро, таким образом, включает в себя 28 журналов. Это можно считать более реалистичной оценкой объема «ядерной», самой спрашиваемой литературы. Отметим, что весь участок, на котором хорошо выполняется закон Брэдфорда при делении на 12 зон, охватывает, как было сказано ранее, журналы с 28-го по 195-й и отрезок на (логарифмической) оси абсцисс [3.3, 5/3]. Середина этого участка находится в точке 4.3, что совпадает с найденной в (1) точкой перегиба аппроксимирующего кубического полинома. Наше предположение о расположении этой точки в середине прямолинейного участка подтвердилось, и два способа оценки величины ядра (оценка сверху и уточненный показатель) находятся в согласии друг с другом.

Заключение

В настоящей работе рассмотрена возможность моделирования процесса использования онлайн-баз данных научной периодики на основе информетрического закона Брэдфорда. Исследован большой массив экспериментальных данных, в результате чего установлено, что модель Брэдфорда в целом может использоваться при анализе данного процесса.

Вид кривой рассеяния спроса на онлайн-статьи по изданиям («библиографа») близок к классическому, что иллюстрирует выполнение графической формулировки закона Брэдфорда. Отклонение библиографа вниз на финальном отрезке (в области малоспрашиваемой литературы) также предсказано в исследованиях закона – эффект ‘Groos Droop’. Как показано, его большая величина и протяженность приводят к тому, что «вербальная», математическая формулировка закона Брэдфорда выполняется в рассматриваемом случае плохо – только на определенном интервале и для достаточно «частого» деления на зоны Брэдфорда. Можно предположить, что причины столь ярко выраженного Groos Droop также совпадают с описанными в литературе: согласно [4] таковой может быть недостаточно полный учет, потеря библиографии в области малопродуктивных по исследуемой теме изданий. Согласно [7] мультидисциплинарность массива исследуемых документов также способна привести к Groos Droop. В нашем случае могут иметь место оба эффекта, если перенести их формулировки от процесса появления журнальных публикаций к их использованию: исследуемое множество онлайн-источников, конечно, не исчерпывает все информационные потребности читателей вуза, и многие периферийные издания, публикующие мало статей, релевантных тематике ГУ-ВШЭ, остаются за рамками исследуемых баз данных и настоящего анализа. С другой стороны, учтенные онлайн-библиотеки существенно политематичны, равно как и интересы, спрос на информацию у читателей вуза широкого профиля, специализирующегося на экономике, менеджменте, социологии, праве, политологии и др. Именно сочетанием этих причин может объясняться особенно сильная деформация библиографа в области малоспрашиваемых изданий.

Применив модель Брэдфорда, удалось определить величину ядра читательского спроса, причем оценка сверху оказалась равной 77 изданиям, более точный показатель еще меньше – 28 журналов. Выделенное ядро обладает неожиданной компактностью – на фоне всего объема электронной подписки (более 13000 наименований) и даже массива изданий, к которым наблюдались обращения (2590 журналов). В данном случае применение математического моделирования при помощи информетрического закона Брэдфорда позволяет сделать практические выводы и сконцентрировать особое внимание при обеспечении информационных нужд аудитории на выделенных «ядерных» изданиях (например, предоставить их как в электронном, так и в печатном виде, заполнить лакуны и восстановить весь архив, избежать прекращения подписки и доступа и т. д.).

Рассмотренные проблемы требуют дальнейшего изучения. Необходимо проверить действие модели Брэдфорда на материале онлайн-статистики различных типов организаций (классических университетов, вузов узкого профиля, научно-исследовательских институтов, публичных библиотек), проанализировать зависимость библиографа от типов ресурсов (в частности, для узкоспециализированных баз данных или ресурсов, предоставляющих доступ к нежурнальной информации). Возможно изучение рассеяния спроса не по наименованиям журналов, а, например, по авторам, издательствам или по пользователям ресурсов. Эти комплексные исследования помогут дать ответ на вопрос о том, можно ли свободно переносить закон Брэдфорда одновременно из «печатной» среды в электронную и от публикационных процессов, для которых он был открыт, на процессы потребления информации, или же потребуется существенная его корректировка и создание отдельного направления информетрии, «информетрии электронного спроса».

Summary

V.V. Pislyakov. Modeling of the electronic information usage on the basis of Bradford informetric law.

Applying of the Bradford informetric law to the modeling of the electronic journals usage is considered. Two versions of the Bradford model are tested on the experimental statistical data. Graph of scattering of the use across titles is plotted and the compact core of the most often used sources is defined..

Литература

1. *Egghe L., Rousseau R.* Introduction to Informetrics: Quantitative Methods in Library, Documentation and Information Science. – Amsterdam: Elsevier Science Publ. 1990. – 450 p.
2. *Bradford S.C.* Sources of information on specific subjects // Engineering. – 1934. – V. 137. – P. 85–86.
3. *Bradford S.C.* Documentation. – London: Crosby Lockwood & Son, 1948. – 156 p.
4. *Brookes B.C.* Bradford's law and the bibliography of science // Nature. – 1969. – V. 224. – P. 953–956.
5. *Groos O.V.* Bradford's law and Keenan-Atherton data // Journal of Documentation. – 1967. – V. 19, No 1. – P. 46.
6. *Qiu L., Tague J.* Complete or incomplete data sets. The Groos Droop investigated // Scientometrics. – 1990. – V. 19, No 3–4. – P. 223–237.
7. *Egghe L., Rousseau R.* Reflections on a deflection: A note on different causes of the Groos Droop // Scientometrics. – 1988. – V. 14, No 5–6. – P. 493–511.
8. *Kendall M.G.* The Bibliography of Operational Research // Operational Research. – 1960. – V. 11, No 1–2. – P. 31–36.
9. *Leimkuhler F.F.* The Bradford Distribution // Journal of Documentation. – 1967. – V. 23. – P. 197–207.
10. *Donohue J.S.* Understanding scientific literatures: a bibliometric approach. – Cambridge: MIT Press, 1973. – 101 p.
11. *Eto H., Candelaria P. M.* Applicability of the Bradford Distribution to International Science and Technology Indicators // Scientometrics. – 1987. – V. 11, No 1–2. – P. 27–42.
12. *Cline G.S.* Application of Bradford's Law to Citation Data // College & Research Libraries. – 1981. – V. 42, No 1. – P. 53–61.

13. *Tonta Y., Al U.* Scatter and obsolescence of journals cited in theses and dissertations of librarianship // *Library & Information Science Research*. – 2006. – V. 28, No 2. – P. 281–296.
14. *Chung Y.* Bradford distribution and core authors in classification systems literature // *Scientometrics*. – 1994. – V. 29, No 2. – P. 253–269.
15. *Stevens S.R.* Mapping the literature of cytotechnology // *Bulletin of the Medical Library Association*. – 2000. – V. 88, No 2. – P. 172–177.
16. *Bulick S.* Book Use as a Bradford-Zipf Phenomenon // *College & Research Libraries*. – 1978. – V. 39. – P. 215–219.
17. *Goffman W., Morris Th.G.* Bradford's Law and Library Acquisitions // *Nature*. – 1970. – V. 226. – P. 922–923.
18. *Tonta Y., Ünal Y.* Scatter of journals and literature obsolescence reflected in document delivery requests // *Journal of the American Society for Information Science and Technology*. – 2005. – V. 56, No 1. – P. 84–94.
19. *Писляков В.В.* Спрос на электронные журналы в университетской библиотеке: работает ли правило Парето? // *Научно-техническая информация. Сер. 1*. – 2005. – № 12. – С. 27–32.
20. *Egghe L.* Applications of the Theory of Bradford's Law to the Calculation of Leimkuhler's Law and to the Completion of Bibliographies // *Journal of the American Society for Information Science*. – 1990. – V. 41, No 7. – P. 469–492.
21. *Drott M.C., Griffith B.C.* An Empirical Examination of Bradford's Law and the Scattering of Scientific Literature // *Journal of the American Society for Information Science*. – 1978. V. 29, No 5. – P. 238–246.

Поступила в редакцию
01.11.07

Писляков Владимир Владимирович – заместитель директора по управлению электронными ресурсами библиотеки Государственного университета – Высшей школы экономики, г. Москва.

E-mail: pislyakov@hse.ru