

СОДЕРЖАНИЕ

Джонсон К. Социальный капитал и исследования по библиотековедению и информатике: хаос в сфере определений или последовательная исследовательская инициатива?	3
Николаисен Й., Франнсен Т. Ф. Фокус-фактор: динамичное измерение специализации журналов	16
Мигель С., Оливейра Э.Ф.Т., Грасио М.К. К. Научная продукция по открытому доступу: широкомасштабный библиометрический анализ в академическом и научном контексте	32
Пош Л. Обогащение онтологий с помощью энциклопедического базового знания для индексирования документов	43

**Главный редактор
БИКТИМИРОВ М.Р.**

**Заместитель главного редактора
ГИЛЯРЕВСКИЙ Р.С.**

**Редакторы:
КОБЗЕВА Л.В.
ОВЧЕНКОВА Е.А.**

Социальный капитал и исследования по библиотековедению и информатике: хаос в сфере определений или последовательная исследовательская инициатива?*

Катрин А. ДЖОНСОН
(Catherine A. JOHNSON)

Факультет изучения информации и медиасредств, Университет Западного Онтарио, г. Лондон, провинция Онтарио, Канада

Статья представляет обзор исследования, проведенного в рамках понятия социальный капитал и опубликованного учеными области библиотековедения и информатики. 99 статей, отвечающих критериям специфической ориентации на библиотековедение и информатику, были выявлены из двух периодических баз данных: Library and information science and technology abstracts и Library literature. Каждая статья прочитывалась с целью определения, какие теоретические области социального капитала цитировались, каковы исследовательские области ученых и как социальный капитал понимался и измерялся. Обзор показал, что хотя социальный капитал понимался до некоторой степени по-разному в зависимости от фокуса исследования, как правило, социальный капитал исследовался или с общественной точки зрения, или индивидуальной точки зрения (а иногда и с той, и с другой) и был определен в качестве выгоды, полученных как от соединяющих (bonding), так и объединяющих (bridging) связей. Хотя понимание социального капитала через различные исследовательские области в библиотековедении и информатике в основном сходится, все еще остаются вопросы относительно измерения обоснованности.

ВВЕДЕНИЕ

Социальный капитал в последние годы стал одной из наиболее динамичных областей исследования в социальных науках. Проявление роста интереса к этой исследовательской области очевидно из поиска в БД *Web of Science*, который открыл ряд статей, упоминающих социальный капитал либо в названии, либо в теме – от 15 статей в 1993 г. до 943 статей в 2014 г., всего 9 512 статей за данный период. Согласно Вулкоку [1, с. 470], можно с уверенностью сказать, что социальный капитал входит в основной поток академической исследовательской инициативы. Другим показателем популярности понятия является изобилие возникших определений. Один ученый подсчитал, что за два последних десятилетия появилось свыше 1 200 определений социального капитала [2, с. 15]. Это отражено в одной из основных

критик социального капитала: «хаос в определении (*definitional chaos*)» [3, с. 5], отсутствие последовательной концептуальной системы и, следовательно, трудность в разработке строгой теории социального капитала. Тем не менее, понятие оказалось популярным в широком профиле академических дисциплин, включая библиотековедение и информатику.

Использование понятия социального капитала в исследованиях по библиотековедению и информатике возникло в 1999 г. вместе со статьей [4], изучавшей влияние вычислительных машин для домашнего пользования на приобретение знаний с применением социального капитала в качестве управляющей переменной. Вторая статья появилась в 2000 г., когда Технологический университет (г. Сидней) [5] опубликовал отчет об исследовании, проведенном среди публичных библиотек Австралии, которые использовали понятие социальный капитал, чтобы узнать, как библиотеки помогают создать доверие в сообществах. Реальным импульсом к интересу относительно изучения пересечения библиотек и социального капитала оказалось выступле-

* Перевод Johnson C.A. Social capital and library and information science research: Definitional chaos or coherent research enterprise. – <http://www.informationr.net/ir/20-4/paper690.html>

ние Роберта Патнэма на встрече Американской библиотечной ассоциации в 2000 г. и включение главы, названной *Branch libraries: The heartbeat of the community* (Отраслевые библиотеки: волнение сообщества) и сфокусированной на том, как социальный капитал генерировался окружающей отраслевой библиотекой в Чикаго, в его книгу *Better together: Restoring the American community* (Лучше вместе: возрождение американского сообщества) [6]. Поиск термина *социальный капитал* в названии, реферате или ключевых словах в БД *Library, information science and technology abstracts* показал, что, начиная с 1998 г., имелось свыше 200 академических журнальных статей, сообщающих либо об исследованиях, которые строили свое изучение с использованием теории социального капитала, либо концептуализировали, как теория социального капитала может использоваться в библиотечном и информационном исследовании. Данная статья изучает упомянутые статьи на предмет определения, имеется ли последовательность в том, как социальный капитал концептуализируется и определяется в библиотечном и информационном исследовании. Она начинается с обзора происхождения понятия социального капитала.

Социальный капитал и анализ социальной сети

Пока концентрация внимания сообщества на социальном капитале определенно затмевает другие исследовательские усилия; понимание происхождения социального капитала в структурной социологии и в частности в теории социальной сети помогает внести ясность в важность взаимоотношений или связей в создании социального капитала. Основное различие между теоретиками социальной сети и теоретиками социального капитала, концентрирующимися на выгодах общества от социального капитала, состоит в том, что первые принимают в качестве единицы исследования взаимоотношения между индивидуумами, а вторые концентрируются на социальном капитале сообщества в целом. Аналитики социальной сети интересуются влиянием социальной структуры на индивидуумов, то есть расположением индивидуумов в структуре и как конфигурация и характеристики связей, которые составляют социальную сеть, влияют на их доступ к ресурсам [7]. Бёрт [8, 9] показал важность социальной структуры через ее идентификацию как *структурных дыр* (structural holes), являющихся пробелами между двумя или более сетями, которые люди должны заполнить и, следовательно, занять посредническую позицию между сетями. Посредник должен не только связать две сети, но также и контролировать идущую между сетями информацию. Согласно Бёрту, «аргумент структурной дыры определяет социальный капитал в терминах преимуществ информации и контроля, являясь посредником в отношениях между людьми, иначе разведенными в социальной структуре» [9, с. 340].

Концепция Грановеттера [10] относительно *силы слабых связей* (strength of weak ties) объясняет выгоды, полученные от различных типов взаимоотношений между индивидуумами. Он разъясняет, что когда люди ищут информацию о новой работе, наиболее выгодными связями являются скорее связи со знакомыми, а не с близкими друзьями. Преимущество слабых связей состоит в том, что они связывают людей в сети, содержащие новую информацию, которую они не могут получить из тесных связей. Следовательно, наличие слабых связей в вашей сети должно быть социальным капиталом, который дает вам преимущество над другими. В обеих концепциях, как Бёрта, так и Грановеттера, относительно значения структурного местоположения и ка-

чества связей, основное предположение заключается в том, что эти связи часто приобретаются с помощью инструментальных средств.

Сильные связи, которые характеризуются доверием и взаимодействием, и слабые, связывающие в новые сети, относятся к соединяющим и объединяющим понятиям, которые составляют основные компоненты теории социального капитала. В то время как объединяющий социальный капитал обычно ассоциируется с индивидуальным уровнем социального капитала, соединяющий социальный капитал больше ассоциируется с точкой зрения сообщества или коллектива на социальный капитал. Соединяющий социальный капитал основан на идее, что большие, плотные сети и люди с общими ценностями могут быть ресурсными друг для друга, работая вместе для достижения взаимных конечных выгод. Он более закрыт по природе и содержится в рамках более сильных отношений между членами группы. Объединяющий социальный капитал включает слабые связи, объединяющиеся в различные социальные сети и таким образом обеспечивающие больший доступ к разнообразным и часто более высокого качества ресурсам.

Эти два различных влияния социального капитала, коллективные и индивидуальные выгоды социальных взаимоотношений и два различных типа социального капитала, объединяющие и соединяющие связи, представляют основные направления исследования социального капитала. В следующем разделе будут рассмотрены статьи, опубликованные в журналах по библиотековедению и информатике, с целью определения, как эти понятия включены в исследование по библиотековедению и информатике.

МЕТОД

Для получения выводов относительно того, как теория социального капитала используется в исследовании по библиотековедению и информатике, были просмотрены ранее упомянутые базы данных. Проводился поиск по термину *социальный капитал* (social capital) в названии, реферате, предмете или в приведенных автором ключевых словах в обеих базах данных до декабря 2014 г. Обнаружена 201 запись статей, опубликованных на английском языке в рецензируемых журналах. В целях дальнейшего усовершенствования поиска были включены только те статьи, которые имели определенную ориентацию на библиотековедение и информатику и фокусировались на социальном капитале как на теоретической основе. Названия журналов также были выверены по списку ядерных журналов по библиотековедению и информатике, которые скомпилированы по обзорам мнений экспертов, темам принятия и распространения, импакт факторам и по тому, содержали ли они статьи локальных факультетов библиотековедения и информатики [11]. Однако не были исключены статьи и из журналов, не входящих в данный список, так как он опускал статьи, написанные учеными и на многих факультетах библиотековедения и информатики или представляющие для них специфический интерес. Например, статьи Гонга, Джемсона и Ченя [12] и Свендсена [13], хотя и неопубликованные в журналах, включенных в ядерный список журналов по библиотековедению и информатике, отвечали другим критериям и имели сильный фокус относительно библиотековедения и информатики (оба журнала касались библиотек). Это сократило перечень до 99 статей, из которых 79 представляли эмпирические исследования, а 20 – концептуальные статьи, редакционные статьи или обзоры иссле-

дования по социальному капиталу. Все статьи были прочитаны, чтобы убедиться, что социальный капитал является основным фокусом статьи, содержащей определение социального капитала, а для эмпирических исследований внимание сосредотачивалось на обзоре литературы и разделах с описанием методологии. Поскольку исследователи сферы библиотековедения и информатики не ограничивались журналами, доступными через поиск в базах данных, предполагается, что найденные статьи обеспечивают репрезентативную выборку типа и сферы исследования по библиотековедению и информатике, использующего социальный капитал в качестве концептуальной основы. Данные о том, когда появились эти статьи, представлены на рис. 1.

Как видно из приведенного выше рисунка, число статей по библиотековедению и информатике, концентрирующихся на исследовании социального капитала, постоянно росло, начиная с появления в 1999 г. первой статьи, а годы с 2012 до 2014 показывают растущее число публикации по сравнению с предыдущими годами.

Это дает возможность предположить, что исследование социального капитала продолжает представлять растущий интерес для ученых области библиотековедения и информатики. Ряд журналов, по сравнению с другими, продемонстрировал большую склонность к опубликованию статей по социальному капиталу: *Library & Information Science Research* (11 статей), *Journal of the Association for Information Science & Technology* (7), *Behaviour & Information Technology* (7) и *The Information Society* (6) были известными распространителями исследования по социальному капиталу.

Чтобы проследить, как социальный капитал концептуализируется в исследованиях по библиотековедению и информатике – или как коллективный, или как индивидуальный ресурс, или как сочетание того и другого, необходимо изучить определения социального капитала, используемые исследователями и учеными, которые они цитируют. Рис. 2 представляет число раз цитирования социального капитала учеными в статьях, включенных в наше исследование.

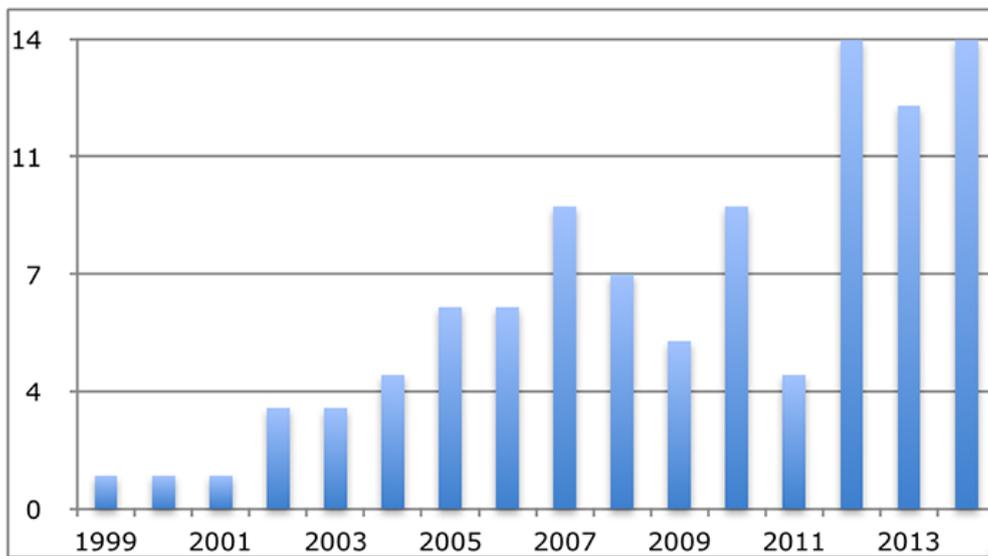


Рис. 1. Число статей по библиотековедению и информатике, опубликованных с 1999 по 2014 г., которые концентрировались на исследовании социального капитала (всего 99 статей)

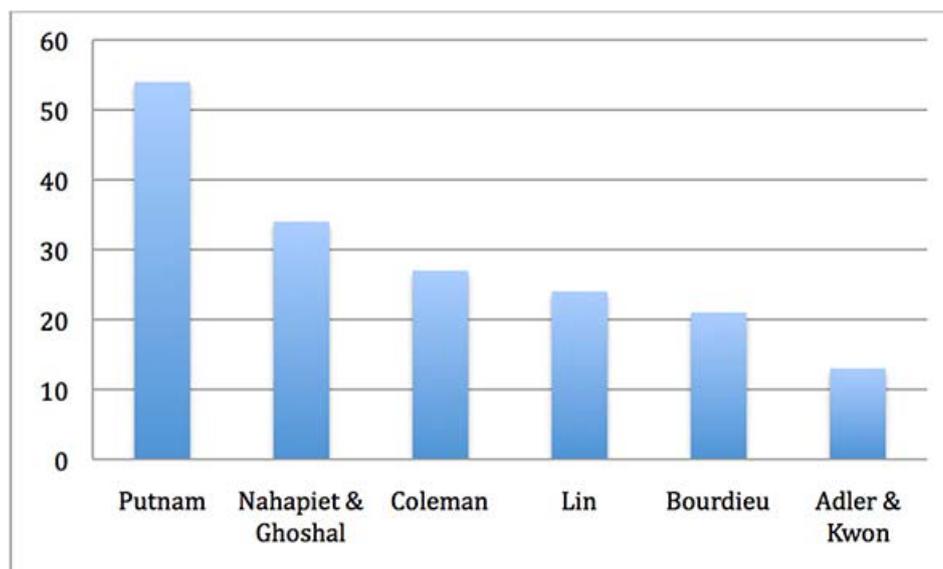


Рис. 2. Наиболее высокоцитируемые ученые, занимающиеся социальным капиталом в исследованиях по библиотековедению и информатике

Обычно самым цитируемым ученым был Роберт Патнэм, который цитировался в 54 статьях. Его книга «Bowling alone: The collapse and revival of the American community» (Игра в кегли в одиночку: крах и возрождение американского сообщества) [14] также является одной из самых цитируемых работ, будучи процитированной свыше 32 тыс. раз, согласно Google Scholar. Коулман с 27 ссылками обычно включается в обзоры литературы вместе с Патнэмом. Типичное определение социального капитала в статьях по библиотековедению и информатике, цитирующих Патнэма, объясняет, что собой представляет социальный капитал, а также в чем его выгоды: социальный капитал «относится к сетям, нормам, доверию и взаимному пониманию, которые связывают вместе членов человеческих сетей и сообществ и позволяют участникам действовать вместе более эффективно для осуществления совместных целей» [15, с. 449]. Большинство статей, фокусирующихся на концептуализации Патнэма, подчеркивают общественные выгоды социального капитала: «Социальный капитал ассоциируется с множественными общественными разработками, демократией, экономическим развитием, эффективностью правительства, развитием сообщества, обучением, здоровьем индивидуума и благосостоянием, с борьбой с преступлениями, с злоупотреблением лекарственными препаратами и подростковой беременностью» [16, с. 878]. Другой пример взят из работы Као, Лю, Донга и Танга, которые объединили определения из работ Нахапет (Nahapiet) и Гошала (Ghoshal) [17] и Патнэма [14] в своем определении социального капитала: «Социальный капитал – это ресурс, который помогает поддерживать сообщество..., социальный капитал стимулирует сотрудничество и кооперацию между членами групп для взаимной выгоды. Таким образом, теория социального капитала может охватывать основное содержание обмена информацией и общественного сотрудничества» [18, с. 1672].

Статья Нахапет и Гошала [17] является одной из самых цитируемых в исследованиях по социальному капиталу (11 106 ссылок согласно базе данных Google Scholar) и также была высоко цитируемой в нашем исследовании (34 статьями из 99). Хотя в основном это представляет интерес для административных ученых, исследователи области библиотековедения и информатики также считают интерпретацию Нахапет и Гошала относительно социального капитала особенно полезной в понимании совместного использования информации внутри или среди организаций. Зная о различном внимании ученых к тому или другому измерению социального капитала (соединяющий или объединяющий капитал), Нахапет и Гошала настаивают на том, что социальный капитал не является понятием одного измерения. В их определение социального капитала входит идея ресурсов в социальных сетях, доступных индивидуумам, а также тесные, создающие доверие связи, которые ведут к совместному использованию информации. Они определяют социальный капитал «как сумму актуальных и потенциальных встроенных ресурсов, доступных через сети и полученных из сетей взаимоотношений, которыми обладает либо индивидуум, либо социальная единица» [17, с. 243]. Они определяют три измерения социального капитала: структурное, относительное и когнитивное. Структурный социальный капитал относится к общей модели связей между действующими лицами: плотность, соединения и иерархия.

На Грановеттера [10], Бёрта [8] и других структурных социологов ссылаются в связи со структурным социальным капиталом. Относительный социальный капитал понимается как активы, созданные и действующие через

взаимоотношения. Эти активы носят скорее поведенческий характер, чем структурный, и состоят из доверия, норм и санкций, обязательств и ожиданий, идентичности и идентификации [17, с. 244]. Основные ученые, которых они цитируют, связаны с этими понятиями и включают Фукуяму [19], Патнэма [20, 21], Бурдье [22], Коулмана [23] и Бёрта [8]. Когнитивное измерение социального капитала – это их уникальный вклад в развитие теории социального капитала и не часто включенный в литературу за пределами исследований по управлению. Когнитивные ресурсы относятся к «ресурсам, обеспечивающим совместные репрезентации, интерпретации и значимые системы среди частей» [17, с. 244]. Это понятие важно для авторов, так как относится к тому, как социальный капитал ведет к созданию интеллектуального капитала, результату обмена совместно используемым знанием и информацией.

Тесно связанной с концептуализацией Нахапет и Гошала является концептуальная модель Адлера и Квона [24], она цитируется тринадцать раз статьями в нашем исследовании, как правило, вместе с Нахапет и Гошалам. Адлер и Квон называют социальный капитал зонтичным понятием, «которое привлекает ученых с точки зрения гетерогенных теоретических перспектив» [24, с. 18]. Их определение совмещает оба понятия социального капитала и констатирует то, как этот капитал определяется и какие понятия подчеркиваются, меняется в зависимости от того, делается ли акцент на сущности, источниках или влияниях социального капитала и включают ли единицы исследования отношения между действующими лицами (объединяющие связи – bridging ties), или структуру отношений в рамках коллектива (соединяющие связи – bonding ties), или и те и другие [24, с. 19]. Их определение содержит внутренние и внешние связи и позволяет социальному капиталу быть отнесенным как к отдельному участнику, так и к коллективному: «Социальный капитал является проявлением доброй воли, доступным отдельным лицам или группам. Его источник лежит в структуре и содержании социальных отношений действующих лиц. Его эффекты воздействия происходят из информации, влияния и солидарности, которые он делает доступными участнику» [24, с. 23]. Их структура социального капитала включает три измерения или аспекта: возможность, мотивация и способность (компетенция), которые хорошо соотносятся с измерениями Нахапет и Гошала. Возможность относиться к структурным аспектам социального капитала: типы связей, частота взаимодействия и количество связей, к которым подсоединены индивидуумы. Мотивация схожа с относительным измерением социального капитала Нахапет и Гошала: нормы, доверие и распределение ценностей среди связей, которые либо могут оставаться в бездействии, но также могут быть инструментально доступными для достижения определенных целей. Компетенция схожа с идеей Нахапет и Гошала относительно когнитивного измерения или общего доверия между членами коллектива. Как и в случае Нахапет и Гошала, это третье измерение не достаточно определено и является трудным, чтобы сделать различие между структурными и относительными формами социального капитала. Способность, определенная как «компетенция и ресурсы в узлах сети» [24, с. 26], также может быть понята как качество ресурсов в сети, которая, согласно Линно [25] и другим теоретикам социальной сети, является основным понятием в рамках теории социального капитала.

С двадцатью четырьмя ссылками в выборке, а также с 6 310 ссылками в соответствии с Google Scholar,

Нань Линь также является важным ученым в сфере исследований социального капитала. Теория социального капитала Линя [25] непосредственно развита из теории социальной сети и теории социальных ресурсов, которые он разработал в 1980-х гг. [26]. Линь строит и расширяет теорию социального капитала на понятиях структурного расположения Бёрта [8] и Грановеттера [10], которая подчеркивает преимущества, полученные от качества социальной сети индивидуума. Сетевая теория социального капитала Линя объясняет, как социальный капитал возникает из структурной сети, и обеспечивает нас большим пониманием казуального направления социального капитала [25, 27, 28]. Определение Линя близко к основанному на классах анализу Бурдье, поскольку он заявляет, что качество социального капитала растет по мере продвижения в более высокие уровни социальной структуры. Линь определяет социальный капитал как «*ресурсы, встроенные в социальные сети, доступные и используемые участниками процесса для действий*» [25, с. 25]. Он рассматривает аккумуляцию слабых (или объединяющих - bridging) связей в качестве обдуманной стратегии для получения доступа к ресурсам более высокого качества, которые часто доступны со стороны людей в более удаленных сетях. Кроме того, люди, которые имеют множество связей, соотносённых с различными навыками и ценностями (разнообразие), проявляют тенденцию делать лучше, чем те, чьи сети в основном состоят из друзей и членов семьи [29].

Линь измеряет социальный капитал используя позиционный генератор, вычисляющий социальный капитал на основе качества связей, к которым индивидуум имеет доступ. Позиционный генератор состоит из перечня профессий или других структурных позиций, которые представляют иерархические позиции в социальной структуре [25, с. 88]. Участников исследования просят указать, знают ли они кого-либо в каждой из этих профессий, чей статус определяется эмпирически полученными оценками престижа. Качество социального капитала основано на комбинации достижения самого высокого по рангу положения, к которому индивидуум имеет доступ; разнообразие, число различных позиций, к которым индивидуум имеет доступ; экстенсивность, разница между самой высокой и самой низкой полученной позицией. Несмотря на концентрацию внимания Линя на инструментальных связях, он также признает важность экспрессивных форм социального капитала, которые являются социальной и эмоциональной поддержкой, полученной от близких друзей и членов семьи. Тем не менее, эти экспрессивные формы социального капитала редко изучаются в исследованиях, проводимых Линем.

Пьер Бурдье, с двадцатью одной ссылкой на него авторами, в нашем исследовании, а также с 22 тыс. ссылок, в соответствии с Google Scholar, на его статью *The forms of capital* (Формы капитала) [22], был наиболее часто цитируемым в обзорах литературы, когда объяснял происхождение понятия социальный капитал, но его идеи, как правило, не развивались далее цитирования его определения социального капитала (см. например, [30, 31, 32]). Однако в ряде случаев концептуализация Бурдье социального капитала была центром изучения. Например, занимаясь распространением понятия социальный капитал за рамками его оригинальной сферы, Юань, Гей и Хембрук [33] ограничили свое исследование до определения Бурдье: «*свойства структуры сети, к которой члены сети могут иметь доступ и активировать, чтобы достичь своих*

целей» [33, с. 27]. Другие ученые концентрировались на анализе на основе классов Бурдье. Исследование Линя и Ченя [34] относительно дифференциальной силы среди теггеров (taggers) было основано на концепции Бурдье, что владение культурным и социальным капиталом определяет чей-то статус и влияние [34, с. 541]. Исследование Моньяру-Папаконстантину и Цапарони [35] относительно образовательных траекторий студентов области библиотекведения и информатики фокусировалось на понятии Бурдье пристрастия, которое «*обеспечивает средства для понимания того, как различия социального, культурного и экономического характера могут влиять на способ, с помощью которого совершеннолетние воспринимают доступные возможности, и привести их к тому, чтобы сделать соответствующий выбор*» [35, с. 240]. Поскольку Нахапет и Гошал [17] также концентрировались на понятии Бурдье ценностей сети, ряд исследований по управлению знанием также ссылался на Бурдье [36-38] при обсуждении относительного измерения социального капитала.

Хотя и индивидуальная, и коллективная выгода должны рассматриваться при оценке социального капитала, многие исследования концентрируются на том или другом, основываясь на контексте и мотивации изучения. В следующем разделе будет рассмотрена литература по библиотекведению и информатике на предмет социального капитала, чтобы определить среди прочего, включают ли изучения оба понятия - соединяющие и объединяющие связи, и использовалось ли одно понятие чаще, чем другое.

СОЦИАЛЬНЫЙ КАПИТАЛ И ИССЛЕДОВАНИЯ ПО БИБЛИОТЕКОВЕДЕНИЮ И ИНФОРМАТИКЕ

Понятие социальный капитал подпадает под несколько областей исследования по библиотекведению и информатике: публичные библиотеки (27%), офлайн и онлайн сообщества (24%), управление знанием (19%), ИКТ – информационные и коммуникационные технологии, использование и влияние на общество (13%), поиск информации (6%), достижения в образовании (4%), обучение онлайн (4%), авторское сотрудничество (3%); одна статья концентрировалась на измерении социального капитала. Число статей в каждой исследовательской области представлено на рис. 3.

Публичные библиотеки и исследование социального капитала

Двадцать одна статья, освещающая эмпирическое исследование, была опубликована, концентрируя внимание на том, как публичные библиотеки помогают создать социальный капитал как в сообществах, где расположены библиотеки, так и для индивидуальных пользователей библиотеки. Хотя более ранние исследования пришли из Австралии, большинство современных исследований в основном поступает из двух источников: *the Public Library – Arenas for Citizenship (PLACE)* - проект, осуществляемый в Норвегии, который дал десять статей, и исследования, проведенные Катрин Джонсон и Метью Гриффисом в США и Канаде, представившие шесть статей.

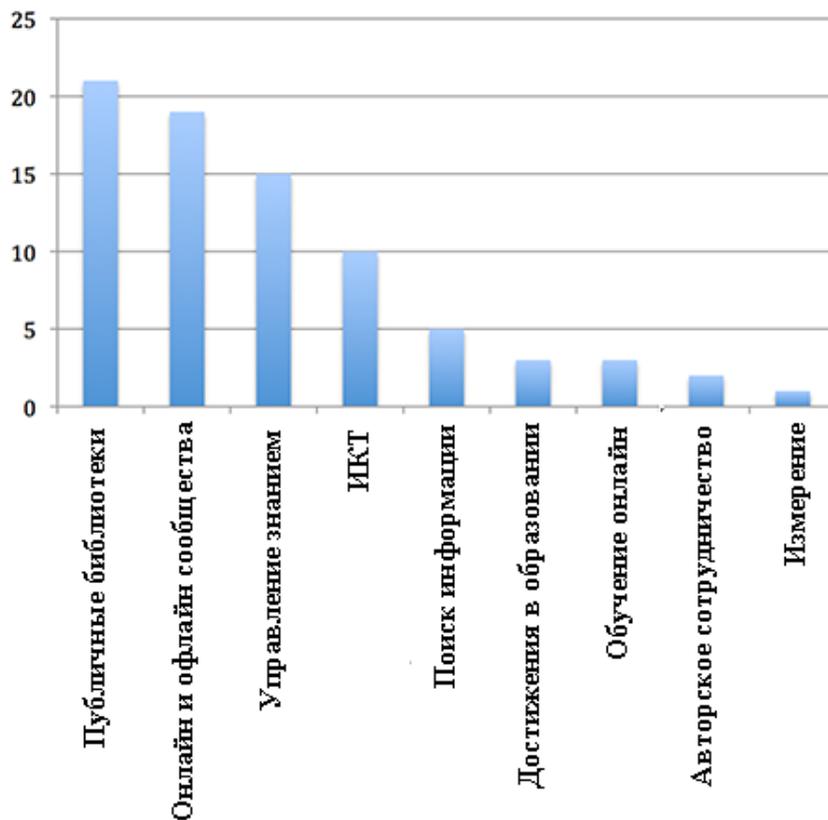


Рис. 3. Исследовательские области библиотековедения и информатики, использующие понятие социальный капитал (на основе эмпирических научных исследований) (n = 79)

Так как публичные библиотеки являются организациями сообществ, понятие Патнэма относительно социального капитала на уровне сообщества используется гораздо чаще для идентификации связей между использованием публичной библиотеки и созданием социального капитала. В исследованиях PLACE социальный капитал обычно интерпретируется как создание всеобщего доверия [39]. Из-за высоких уровней иммиграции в Норвегии ученые были мотивированы на то, чтобы показать, как публичные библиотеки могут помочь группам иммигрантов интегрироваться в норвежское сообщество. Акцент главным образом делается на определение Патнэма, которое включает как соединяющий капитал, так и объединяющий капитал: «В исследовании социального капитала существуют две формы социального капитала: соединяющий социальный капитал в форме сетей и полного доверия (исключительное доверие) между членами компактных и высоко интегрированных групп и объединяющий социальный капитал в форме сетей и неубедительного, слабого доверия (всеобщее доверие) через первичные службы» [40].

Ученые называют это общественной перспективой. Они добавляют еще одну перспективу, называемую институциональной, которая относится к доверию, создающему «эффекты универсальных и беспристрастных публичных учреждений и демократий» [41, с. 259]. Исследователи демонстрируют, что библиотеки полезны для того, чтобы собирать вместе в корне различных людей в безопасном месте, где они имеют возможность узнать друг друга или, по крайней мере, наблюдать друг друга, а следовательно, увеличивать уровни доверия между

ними. Они развили два понятия, связанные с объединяющим и соединяющим социальным капиталом, которые помогают объяснить этот процесс: низкоинтенсивные и высокоинтенсивные места встреч. Низкоинтенсивные места встреч – это места, где люди подвергаются действию «других ценностей и интересов в отличие от тех, которых они сами придерживаются» [42, с. 17]. Эти места, как можно предположить, состоят из потенциальных или латентных, слабых объединяющих связей. Высокоинтенсивные места встреч представляют совсем противоположное – это места, где собираются люди, которые, как правило, разделяют схожие ценности, привычки и мировоззрение, которые эквивалентны соединяющим связям.

Пытаясь понять, откуда происходит социальный капитал, ученые PLACE строят гипотезу, что первым шагом к созданию доверия (считается эквивалентным социальному капиталу) было сократить неравенство. Поскольку библиотеки представляют собой универсальные учреждения, открытые каждому и обеспечивающие стандартную услугу, то они являются идеальным местом, где можно наблюдать микромеханизмы, которые ведут к большому доверию [43]. Ожидалось, что контакт в библиотеке между новыми иммигрантами и долгое время живущими резидентами станет результатом в развитии доверия между ними и, таким образом, большей интеграции в доминирующем обществе. Однако исследования ученых имеют смешанные результаты. На основе интервью и наблюдений в рамках библиотек исследователи обнаружили, что как высокоинтенсивные, так и низкоинтенсивные встречи проходили в библиотеке

[43, 40, 42] и что большинство взаимодействий, которые имели там место, происходило между друзьями и родственниками и таким образом эквивалентно соединяющему социальному капиталу. Исследование [44], которое включало интервью с девятью иммигрантами-женщинами показало, что через использование библиотеки они установили контакты с большинством населения и таким образом увеличили чувства доверия по отношению к ним.

Схожее исследование Вархейма, на этот раз включающее незарегистрированных иммигрантов в Колорадо, показало, что хотя доверие в библиотеке возросло с помощью посещения библиотечных программ, оно не распространилось далее библиотекарей и инструкторов, представляющих программы [41, с. 272]. В этих исследованиях объединяющие связи совсем необязательно обеспечивают доступ к новым ресурсам, как в определении Линя, они скорее представляют возможный вход в доминирующее общество через развитие доверия. Согласно авторам, формирование объединяющих связей или всеобщего доверия является первым шагом к созданию сильных, интегрированных сообществ.

Исследования, проведенные Джонсон [45-47], Джонсон и Гриффисом [48, 49] и Гриффисом и Джонсон [50], посмотрели на влияния использования библиотек как на индивидуальный социальный капитал, так и на социальный капитал сообщества. Эти исследования применяли и количественные, и качественные научные методы, а также комбинированные подходы Патнэма и Линя для обеспечения более целостного взгляда на важность библиотек для индивидуума и сообщества. Соединяющий социальный капитал измерялся с помощью использования анкет, разработанных Патнэмом на основе *SK 2006 Community Benchmark Surveys* [51], которые измеряли социальный капитал на основе уровней вовлечения сообщества, гражданских обязательств и доверия. Анкеты отправлялись пользователям как городских, так и сельских библиотек. Исследователи обнаружили значительные статистические ассоциации (сходства) между измерениями социального капитала в городских, но не в сельских библиотеках [46, 48, 49]. Объяснение, приведенное для этого различия, состояло в том, что библиотеки небольших городов в основном часто посещались представителями среднего класса, уже имеющими высокие уровни доверия, а городские библиотеки имели более высокую долю бедных посетителей, которые получали выгоду от социальных и информационных ресурсов, доступных через библиотеку [49, с. 187]. Объединяющий социальный капитал измерялся с помощью позиционного генератора Линя [25]. Негативные ассоциации были найдены между индивидуальными измерениями социального капитала и использованием библиотекой в городских пространствах, полагая, что люди совсем необязательно увеличивают свой уровень социального капитала путем использования библиотеки, а скорее использование библиотеки индивидуумами являлось стратегией, чтобы компенсировать их низкие уровни социального капитала.

Интервью относительно качества также раскрыли взаимодействия, встречающиеся в городских библиотеках, которые являлись предсказателями роста социального капитала, такие как больший доступ к информационным ресурсам в библиотеке, а также к ресурсам вне рамок библиотеки, персональное выполнение обязательств между пользователями и штатом библиотеки, такие как присмотр за детьми, предоставленными в библиотеке самим себе, или предложение специальных программ, проводящихся после школьных занятий [46, 48, 50].

Гриффис и Джонсон [50] также определили четыре основных роли публичной библиотеки, которые должны помочь увеличить социальный капитал в сельских сообществах, включая следующие: работа в качестве как информационного, так и социального центра для общества, помощь в интеграции в общество за счет открытия клубов любителей книг и введения программ для детей через значимость самой физической существующей библиотеки, которая выступает в роли важного источника идентичности сообщества и, наконец, библиотечного обслуживания как одной из частей крупной сети групп сообществ и организаций.

Социальный капитал и онлайн и офлайн сообщества

Исследования сообществ оказались особенно сильной стороной исследования, использующей понятие социального капитала. Многие из них распространяются от ранних исследований сообществ, проведенных социологом Барри Веллманом и его коллегами в 1970-х гг. и позже. В этих исследованиях Веллман показал, что несмотря на тот факт, что люди больше не были тесно связаны в рамках физически объединенных сообществ, они могли поддерживать сильные межличностные связи через использование технологий, таких как телефон и автомобиль [52]. С появлением Интернета ученые стали интересоваться, будет ли такая новейшая технология в итоге разрывать даже эти социальные связи, поскольку люди все больше и больше времени проводили в режиме онлайн. Исследования по библиотековедению и информатике фокусировались как на влияниях использования Интернета на взаимоотношения офлайн, так и на влиянии вовлечения человека в социальные медиа на индивидуальный социальный капитал. Одной из первых статей, рассматривающих этот вопрос, была статья Веллмана, Кван-Хааса, Витте и Хэмптона [53], которая привела свой аргумент в качестве реакции на исследования Патнэма, обнаружившие, что американский социальный капитал снизился после Второй мировой войны [14]. Веллман и др. объясняли, что Патнэм, возможно, смотрел на старомодные способы создания сообщества. Они полагают, что новые пути создания сообщества, через вовлечение в онлайн сообщества, могут на самом деле заменить социальный капитал, который, как заявляет Патнэм, был потерян [53, с. 437]. Они описывают три формы социального капитала: «Сетевой капитал: отношения с друзьями, соседями, родственниками и коллегами, которые обеспечивают общение, эмоциональную помощь, товары и услуги, информацию и чувство принадлежности. Участвующий капитал: вовлечение в политику и добровольные организации, которые предоставляют людям возможности соединять и создавать совместные достижения. Обязательство сообщества: сильное чувство вовлечения сообщества, которое мобилизует их социальный капитал более эффективно» [53, с. 437].

Первые две формы имеют отношение к работе Патнэма, тогда как третья взята из исследований авторов относительно сообщества. Однако во всех этих формах внимание концентрируется на сильных соединяющих связях, обеспечивающих социальную и эмоциональную поддержку и ведущих к чувствам благополучия и доверия. Авторы обнаружили, что Интернет не сокращает социальный капитал, а является еще одним средством, через которое люди могут формировать и поддерживать взаимоотношения, скорее дополняя социальный капитал, чем уменьшая. Однако они также думают, что растущее вовлечение в деятель-

ность Интернета снижает восприятие индивидуумом сообщества в режиме онлайн [52, с. 451].

Основываясь на этом исследовании, Эллисон, Стейнфилд и Лампе [54] изучали пользователей Facebook, чтобы определить, какой была связь между социальным капиталом онлайн и офлайн. Предыдущее исследование показало, что взаимосвязи, сформированные онлайн, сдвинулись в офлайн и, следовательно, увеличат весь общий банк социального капитала индивидуумов [55, 56]. Исследование Эллисона и др. [54] было проведено среди студентов старших курсов – пользователей Facebook – для определения, до какой степени используются сайты социальной сети, чтобы создать новые дружеские отношения или поддерживать старые, используя соединяющие и объединяющие понятия социального капитала. Они выяснили, что провоцируемые Интернетом взаимосвязи приводят в результате к созданию как слабых связей, так и связей в новой сети, в то же время помогающих в поддержке их тесных или соединяющих связей. Ан [57] остановилась на понятиях как соединения, так и объединения, чтобы изучить взаимоотношения между использованием сайтов социальной сети, таких как Facebook и MySpace, и наличием социального капитала среди подростков. Ан использовала работу “*Internet social capital scales*” (Масштабы социального капитала Интернета) Уильямса [58] для измерения взаимоотношения между социальным капиталом и использованием сайта. Соединяющий социальный капитал измерялся путем опроса студентов о существовании (или отсутствии) доверительных взаимоотношений и людей, к которым бы они могли обратиться за советом [57, с. 102]. Объединяющий социальный капитал измерялся путем постановки вопроса относительно того, ведет ли взаимодействие с людьми на Facebook к участникам, становящимся заинтересованными в делах, происходящих за пределами их города, или желающих попробовать что-то новое [57, с. 102]. Она выявила, что существование в качестве члена сайта социальной сети ассоциируется с развитием обоих видов социального капитала (соединяющего и объединяющего), хотя результаты были разнородными. Как заявляет Ан, важность этого исследования заключается в том, чтобы показать родителям и другим опекунам, что использование сайтов социальной сети совсем необязательно имеет вредные влияния, такие как киберзапугивание или напрасная трата времени, а может привести к позитивным социальным взаимодействиям [57, с. 107].

Недавно Апшель и др. [59] обратили внимание на проблемы, связанные со шкалой Уильямса, используемой для измерения социального капитала в онлайн сообществах. Сравнивая инструменты давно установленного и проверенного измерения социального капитала сети, например, генераторы имен [9], генераторы позиций [25] и генераторы ресурсов [60] со шкалой Уильямса [58], они обнаружили, что не было прямой корреляции между полученными данными этих двух методов для измерения социального капитала [59, с. 408]. Это иллюстрирует трудность, которую многие имеют в отношении с теорией социального капитала, так как нельзя быть уверенным, что измеряются те же самые показатели социального капитала.

Еще один пример, показывающий, как может быть построено сообщество через участие в онлайн сообществах, – это исследование людей, страдающих заболеваниями двигательного нейрона, проведенное Лоуном и Д’Алессандро [61]. Исследование использует определение Линя [25] как инструментальных, так и экспрессив-

ных форм социального капитала: ресурсы, встроенные в сети, доступные участникам для действий, с ресурсами, включающими как информационную, так и социальную поддержку [61, с. 168]. Аспект соединения онлайн сообществ представляет собой специфическую тему или интерес, вокруг которых строится сообщество, а объединяющий аспект – это объединение вместе «*географически различающихся индивидуумов из многообразия расовых и образовательных формирований*» [61, с. 168]. Методология для определения наличия социального капитала состояла из идентификации случаев в онлайн разговорах, которые включают примеры создания доверия и общего взаимодействия. Члены онлайн сообщества также выигрывали от многообразия экспертных оценок, присутствующих в слабых связях, включающих группу, которая стала результатом высокополезного потока информации, доходящего до членов. Это прекрасный пример того, как и соединяющие связи (порожденные совместным общим интересом), и объединяющие связи (слабые связи, соединяющие в различные социальные сети) помогают построить сильное сообщество вокруг общего интереса.

Социальный капитал и исследование управления знанием

Статьи Нахапет и Гошала [17] и Адлера и Квона [24] по управлению знанием подтолкнули ряд ученых области информатики к тому, чтобы взглянуть на влияние социального капитала на общее знание или информацию в организациях. Статья Гюйсмана и Вульфа [62] показала, что формальные средства продвижения совместного использования знания, которые отделены от социальной среды, работают недостаточно хорошо. Поэтому организации должны признать важную роль, которую социальный капитал играет в стимулировании служащих к совместному использованию информации. Гюйсман и Вульф утверждают, что «*тщательное исследование достоинства социального капитала сообществ и улучшение уровня средств социального капитала для совместного использования знания будет, вероятно, больше соответствовать возможностям людей, мотивации и способности совместно использовать знание*» [62, с. 44].

Ряд других статей фокусировались на том, как социальный капитал стимулирует совместное использование знания в рамках организаций [63, 64], повышает карьерный рост специалистов в области информационных технологий [65], помогает разработчикам программного обеспечения, занятым в сфере развития программного обеспечения на условиях оффшорного аутсорсинга [36], и влияет на сотрудничество в проектах [66]. Каждая из этих статей объясняла совместное использование знания в организациях на основе моделей, предложенных либо Нахапет и Гошалом [17], либо Адлером и Квоном [24]. Эти исследования относительно организации представили оптимальную ситуацию, где социальный капитал дает выгоды как организации (социальный капитал на уровне сообщества), так и индивидуумам, показывая возможность объединения этих различных измерений социального капитала в одну согласованную теорию.

Социальный капитал и информационно-коммуникационные технологии (ИКТ)

Понятие социальный капитал также использовалось в исследованиях, изучающих цифровое деление, принятие информационных и коммуникационных техноло-

гий как индивидуумами, так и сообществом информатиков (community informatics). Ранее, в 2000-х гг., Варшпауер [67] открыл связь между социальным капиталом и цифровым делением. Он предвидел, что оба уровня – индивидуальный и коллективный – социального капитала должны влиять на способность людей принять информационные технологии. Он предполагал, что как доверие (соединяющие связи), так и слабые связи со знающими людьми в центрах доступа сообщества важны в желании и способности людей использовать эти технологии [67, с. 317-318]. Исследование Уильямса [68] подтверждает важность соединяющего социального капитала в использовании ИКТ, отмеченных уровнями создаваемого между людьми доверия, которые побуждают их принять новые технологии. Основываясь главным образом на концепции социального капитала Линия [25], Чень [69] концентрируется на социальном капитале индивидуального уровня и степени получения людьми помощи от семьи и друзей, чтобы разрешить технические проблемы. Во многих этих исследованиях было обнаружено, что и соединяющие, и объединяющие формы социального капитала имеют влияние на принятие индивидуумами новых информационных технологий.

Сообщество информатиков или связь между ИКТ и развитием сообщества, заинтересовано в определении влияния этих технологий на сообщества, а также мотивации людей на их принятие. Фарук, Жано, Ксяо, Меркель, Россон и Керрол [70] изучали, как совместная разработка сетевого сайта сообщества использовала социальный капитал сообщества в форме слабых и сильных связей с правильными навыками, чтобы совершенствовать сетевой сайт. Миньон и Хенли [71] представили исследования пяти случаев, изучающих влияние внедрения ИКТ на сообщества коренных жителей Канады. Они выявили, что инвестиции отдельных сообществ в технологии могут повысить их объединяющие возможности относительно связи с другими сообществами аборигенов, а также контакты с правительственными и неправительственными учреждениями. Предполагалось, что совместное использование информации среди различных сообществ также повысит их возможность более успешно вести переговоры в отношении прав на владение землей. Схожее исследование, проведенное Ю. Веньжи и Шенглонгом [72], использовало пример методологии для изучения влияния таких технологий на деревни в тибетском регионе Китая. Они обнаружили, что объединяющий социальный капитал имел место в финансовых, образовательных и людских ресурсах, взятых для развития проекта и для того, чтобы сделать его полезным для сельских жителей, и что соединяющий социальный капитал стимулировал людей использовать технологии и таким образом сократить цифровую бедность.

Социальный капитал и поиск информации

Исследования, концентрирующиеся на поведении во время поиска информации, пытаются показать влияние социального капитала на уровне индивидуума на их (индивидуумов) способности найти полезную информацию. Джонсон [73, 45] использовала структуру социального капитала, чтобы исследовать поведение во время поиска информации городского населения Монголии. Рассматривая социальный капитал в качестве ресурсов, встроенных в сети, доступные индивидуумам для их выгоды, исследование основывалось только на определении социального капитала Линем. Члены индивидуальных социальных сетей были выявлены с помощью

использования генераторов имен, которые определяли как близкие, так и удаленные связи на основе запросов относительно имен людей, обеспечивающих определенные уровни помощи. Позиционный генератор Линия [25] также использовался для определения качества ресурсов в сети на основе доступа к людям, размещенным в рамках профессий. Исследование показало, что люди с высоким уровнем социального капитала более вероятно должны найти самые лучшие источники информации, которые рассматривают их проблемы [45], и что они применяли слабые связи, когда использовали людей в качестве источников информации [73].

Исследование Ваудстра, ван ден Хофа и Скутена [74], фокусировавшееся на поиске информации, использовало модель, разработанную Нахапет и Гошлом [17], чтобы определить, как на выбор информационного источника влияло его качество и доступность. В этом исследовании сценарии разрабатывались и представлялись членам бизнес-организации, чтобы определить, к кому надо обращаться за информацией. Они выявляли имена информационных источников, а также влияние качества и доступности источника на их выбор. Качество рассматривалось как имеющее более сильное влияние на выбор источника, чем доступность. Что касается исследования Джонсон [73], то индивидуумы хотели выйти за пределы их зоны комфорта (их тесные связи), чтобы получить информацию лучшего качества, бросая таким образом вызов общему мнению в литературе по библиотековедению и информатике, что выбор человека в качестве источника информации обычно является опцией, требующей меньшего усилия [75, с. 142].

В то время как эти исследования информационного поведения показали важность слабых или объединяющих связей в поиске информации, другие исследования выявили, что соединяющие связи также важны для принятия информации. Вейнот [76, 77] изучала сети пациентов с заболеваниями ВИЧ/СПИД, живущих в сельских местностях Канады, чтобы определить, как они создают информацию или улучшают сети с целью помочь им справиться с этими серьезными заболеваниями. Через интервью с пациентами, страдающими ВИЧ/СПИД, где она установила генератор имен, чтобы выявить прошлых и настоящих членов их сетей, Вейнот смогла продемонстрировать динамичный характер социальных сетей и процесса поиска информации. Она обнаружила, что пациенты с заболеваниями ВИЧ/СПИД лишали себя определенных членов сети, которые более критичны к их ситуации, но добавили новых, состоящих из тесных связей, которые могут обеспечить эмоциональную поддержку, а также профессиональные связи, которые обеспечат информацию хорошего качества. Что касается исследования Лоуна и Д'Алессандро [61], то определение как соединяющих, так и объединяющих связей было важным в понимании всей сферы того, как социальный капитал приносит выгоду индивидуумам.

ОБСУЖДЕНИЕ

Несмотря на заявление Файна [3] относительно «*хаоса в определении*» понятия социальный капитал, большинство ученых области библиотековедения и информатики основывает их концептуализации на работе главных теоретиков социального капитала, таких как Бурдьё, Коулман, Патнэм и Линь. В зависимости от фокуса их исследования они рассматривали социальный капитал либо с точки зрения сообщества, либо с точки зрения индивидуума, а иногда с точки зрения и того, и другого. Они также часто включают в свой анализ и соединяю-

щие, и объединяющие формы социального капитала. Эти две точки зрения, кажется, являются в корне различными способами рассмотрения социального капитала – на уровне сообщества теоретики акцентируют доверие, которое создается через компактные связи социального взаимодействия, а теоретики социальной сети концентрируются на выгодах для индивидуумов от того, что они имеют высококачественные связи в рамках их сетей или слабые связи, которые связывают с лучшими ресурсами. Однако обе эти точки зрения приводят к схожему результату – это выгоды от социальных взаимоотношений для индивидуумов, сообществ и фирм.

Ученые, интересующиеся выгодами от социального капитала для общества или сообщества, обычно следуют подходу Патнэма, который измеряет социальный капитал как наличие доверия, норм участия и значений сильных соединяющих связей. Исследование норвежских ученых, а также Джонсон и Гриффиса демонстрируют, что наличие открытого для всех места и где могут собираться различные члены общества, имеет положительные результаты в создании как осведомленности относительно других членов сообщества, так и в создании через их взаимодействия больших уровней доверия. Подход Линя, фокусирующийся на индивидуальном социальном капитале, хорошо подходит для исследования поиска информации, которое охватывает часто намеренную попытку приобрести полезные контакты. Модель социального капитала Нахапет и Гошала [17] сочетает обе точки зрения: концептуализация Патнэмом доверия, которое приводит к совместному использованию информации, и подход Линя относительно социальной сети, концентрирующийся на социальном капитале, который существует во взаимоотношениях и, будучи доступным, проявляется в лучших результатах для организации.

Сетевые аналитики часто критикуют теоретиков социального капитала на уровне сообщества из-за трудности в определении каузального направления. Их аргумент состоит в том, что при измерении успеха или неудачи действия сообщества на основе наличия или отсутствия социального капитала детерминанты и последствия действия становятся тем же самым, что приводит к замкнутой логике: социальный капитал порождает социальный капитал. Эта замкнутость причины и следствия, как заявляет Портес [7, с. 4], ликвидируется с помощью сетевых аналитиков. Например, подход Линя разделяет причину и следствие, предполагая, что наличие высоких уровней социального капитала, основанное на эмпирически полученных измерениях, выражается в лучших результатах для индивидуума [25, с. 28]. Апфель и др. [59] утверждают, что только с помощью «отделения структурного компонента – социальный капитал – от его результата и вкладчиков, можно будет говорить о затратах и выгодах социальных сетей» [59, с. 399]. Однако для исследователей сообщества логическая замкнутость может на самом деле быть определяющей характеристикой социального капитала сообщества; это ценность (активы), которая создается через повторяющиеся взаимодействия среди людей, приводя к все возрастающей осведомленности и доверию и которая дает возможность индивидуумам и сообществам работать вместе с целью совместного использования информации, а также разрешения общих проблем. Хотя можно приводить доводы, что более целостный взгляд на социальный капитал будет включать обе точки зрения, невозможно изобрести один единственный метод измерения социального капитала, который охватывает как индивидуальные, так и

общественные влияния и является подходящим для всех исследовательских задач.

Несмотря на то, что ученые сферы библиотековедения и информатики в общем соглашаются по вопросу о том, из чего состоит социальный капитал, т.е. возможность работать вместе для достижения общих целей, развитие доверия, позволяющего совместное использование информации, чувство сообщества, ресурсы хорошего качества, доступные через взаимосвязи, но способности, которыми эти влияния измеряются, еще не конвергированы. Аналитики социальной сети и исследователи управления знанием, вероятно, показывают самое большое постоянство в их выборе измерений. Интересно, что хотя концептуализации Патнэма относительно социального капитала, как правило, следует большинство концентрирующихся на сообществе исследований, в области библиотековедения и информатики только незначительное число ученых использует показатели, которые он (Патнэм) разработал для своего обзора “*Social capital community benchmark survey*” [51]. Вместо этого ряд ученых использует доверие в качестве замены для социального капитала, часто измеряемого через ответы на вопросы о доверии в обзорах “*World values survey*” и “*General social survey*”: «Вообще говоря, можете ли вы сказать, что большинству людей можно доверять или что вам не надо быть слишком осторожными в отношениях с людьми» (см. например, [16] и [12]). Другие задают вопрос, может ли социальный капитал измеряться только одним показателем [78] и является ли сам вопрос надежной мерой доверия [79]. Основная проблема в отношении социального капитала может быть в его интуитивной привлекательности; инстинктивно мы знаем, что тот, кого мы знаем, имеет значение, и что социальная сеть хорошего качества важна для благополучия сообществ и индивидуумов. Однако некоторые показатели социального капитала, включая доверие, нормы и взаимность, могут быть слишком слабыми, чтобы охватить поддающиеся управлению переменные.

ЗАКЛЮЧЕНИЕ

Целью данного исследования было выявить, как социальный капитал концептуализируется и определяется в исследовании по библиотековедению и информатике. Для осуществления этого были определены 99 статей через поиск в базах данных периодики по библиотековедению и информатике. Роберт Патнэм, фокусирующийся на общественных выгодах социального капитала, был самым цитируемым ученым в области социального капитала, за ним следовали ученые из сферы управления знанием – Нахапет и Гошала, а также аналитик социальной сети Нань Линь. Социальный теоретик Пьер Бурдьё, который в 1980-х гг. дал толчок к интересу относительно исследования социального капитала, также цитировался высоко. Хотя ученые области библиотековедения и информатики использовали многочисленные формулировки понятия социального капитала, большинство было ограничено определениями, данными этими известными учеными. Понятие социального капитала используется, чтобы объяснить следующее:

- как онлайн сообщества и публичные библиотеки помогают создать и поддерживать социальный капитал [57, 42];
- как информация совместно используется внутри и среди организаций [62, 66];
- как социальный капитал усиливается или насколько он необходим для успешного внедрения проектов по ИКТ и принятия технологий индивидуумами [72, 68];

- факторы в социальных взаимоотношениях, которые приводят к успешным стратегиям поиска информации [73, 77].

Различные способы, в которых понятие интерпретируется и используется, хотя и часто варьируется между точками зрения сообщества и индивидуума, отражают возрастающую согласованную исследовательскую инициативу, которая демонстрирует как эволюционирующий характер понятия, так и его гибкость, чтобы быть релевантным различным аспектам исследования по библиотековедению и информатике. Тем не менее, остаются проблемы относительно надежности измерения, которые следует рассмотреть до того, как мы сможем говорить об одной всеохватывающей теории социального капитала.

Благодарность. Автор выражает признательность анонимным рецензентам и редакторам журнала *Information Research* за их очень полезные замечания и рекомендации, которые значительно улучшили первоначальную рукопись данной статьи.

ЛИТЕРАТУРА

1. *Woolcock M.* The rise and routinization of social capital: 1988-2008// *Annual Review of Political Science*. — 2010. — Vol.13. — P. 469-487.
2. *Dale A., Onyx J. (Eds.)*. A dynamic balance: Social capital and sustainable community development. — Vancouver and Toronto: UBC Press, 2005.
3. *Fine B.* Theories of social capital: Researchers behaving badly. — London: Pluto Press, 2010.
4. *Attenwell P., Battle J.* Home computers and school performance// *Information Society*. — 1999. — Vol. 15, No.1. — P. 1-10.
5. *University of Technology, Sydney and State Library of New South Wales.* 'A safe place to go': Libraries and social capital. — Sydney, NSW: University of Technology, 2000. — http://www.sl.nsw.gov.au/services/public_libraries/docs/safe_place.pdf. (Archived by WebCite at <http://www.webcitation.org/6cNXcOmpM>).
6. *Putnam R. D., Feldstein L.M.* Better together: Restoring the American community. — New York, NY: Simon and Schuster, 2003.
7. *Portes A.* The two meanings of social capital// *Sociological Forum*. — 2000. — Vol. 15, No.1. — P. 1-12.
8. *Burt R.S.* Structural holes. — Cambridge MA: Harvard University Press, 1992.
9. *Burt R.S.* The contingent value of social capital// *Administrative Science Quarterly*. — 1997. — Vol. 42, No.2. — P. 339-365.
10. *Granovetter M. S.* The strength of weak ties// *American Journal of Sociology*. — 1973. — Vol. 78, No.6. — P. 1360-1380.
11. *Nixon J.M.* Core journals in library and information science: Developing a methodology for ranking LIS journals// *College & Research Libraries*. — 2014. — Vol. 75, No.1. — P. 66-90.
12. *Gong H., Japzon A.C., Chen C.* Public libraries and social capital in three New York City neighbourhoods// *Tijdschrift voor Economische en Sociale Geografie*. — 2008. — Vol. 99, No. 1. — P. 65-83.
13. *Svendsen G.L.H.* Public libraries as breeding grounds for bonding, bridging and institutional social capital: The case of branch libraries in rural Denmark// *Sociologia Ruralis*. — 2013. — Vol. 53, No.1. — P. 52-73.
14. *Putnam R. D.* Bowling alone: The collapse and revival of the American community. — New York, NY: Simon & Schuster, 2000.
15. *Widen-Wulff G., Ginman M.* Explaining knowledge sharing in organizations through the dimensions of social capital// *Journal of Information Science*. — 2004. — Vol. 30, No.5. — P. 448-458.
16. *Varheim A., Steinmo S., Ide E.* Do libraries matter? Public libraries and the creation of social capital// *Journal of Documentation*. — 2008. — Vol. 64, No.6. — P. 877-892
17. *Nahapiet J., Ghoshal S.* Social capital, intellectual capital and the organizational advantage// *The Academy of Management Review*. — 1998. — Vol. 23, No.2. — P. 242-266.
18. *Cao Q., Lu Y., Dong D., Tang Z., Li Y.* The roles of bridging and bonding in social media communities// *Journal of the American Society for Information Science & Technology*. — 2013. — Vol. 64, No. 8. — P. 1671-1681.
19. *Fukuyama F.* Trust: Social virtues and the creation of prosperity. — London: Hamish Hamilton, 1995.
20. *Putnam R.D.* The prosperous community: Social capital and public life// *American Prospect*.—1993. — Vol. 13. — P. 35- 42.
21. *Putnam R.D.* Bowling alone: America's declining social capital// *Journal of Democracy*. — 1995. — Vol. 6, No.1. — P. 65-78.
22. *Bourdieu P.* The forms of capital/ J. G. Richardson, (Ed.), *Handbook of theory and research*. (pp. 241-258). — New York, NY: Greenwood Press, 1986.
23. *Coleman J. S.* Foundations of social theory. — Cambridge, MA: Belknap Press of Harvard University Press, 1990.
24. *Adler P.S., Kwon S.-W.* Social capital: Prospects for a new concept// *Academy of Management Review*. — 2002. — Vol. 27, No.1. — P. 17-40.
25. *Lin N.* Social capital: A theory of social structure and action. — Cambridge: Cambridge University Press, 2001.
26. *Lin N., Dumin M.* Access to occupations through social ties// *Social Networks*. — 1986. — Vol. 8, No.4. — P. 365-385.
27. *Lin N., Cook K.S., Burt R.S. (Eds.)*. Social capital: Theory and research. — New York, NY: Aline de Gruyter, 2001.
28. *Lin N., Erickson B.H. (Eds.)*. Social capital: An international research program. — New York, NY: Oxford University Press, 2008.
29. *Erickson B.H.* Social networks: The value of variety// *Contexts*. — 2003. — Vol. 2, No.1. — P. 25-31.
30. *Elbeshausen H., Skov P.* Public libraries in a multicultural space: A case of integration processes in local communities// *New Library World*. — 2004. — Vol. 105, No.3/4. — P. 131-141.
31. *Liao S., Chou E-Y.* Intention to adopt knowledge through virtual communities: Posters vs. lurkers// *Online Information Review*. — 2012. — Vol. 36, No.3. — P. 442-461.

32. *Oztoke M.* Tacit knowledge in online learning: Community, identity, and social capital// *Technology, Pedagogy & Education*. — 2013. — Vol. 22, No.1. — P. 21-36.
33. *Yuan Y.C., Gay G., Hembrooke H.* Focused activities and the development of social capital in a distributed learning community// *The Information Society*. — 2006. — Vol. 22, No.1. — P. 25-39
34. *Lin C.-S., Chen Y.-F.* Examining social tagging behaviour and the construction of an online folksonomy from the perspectives of cultural and social capital// *Journal of Information Science*. — 2012. — Vol. 38, No.6. — P. 540-547.
35. *Moniarou-Papaconstantinou V., Tsatsaroni A.* Educational trajectories of LIS students: Continuities and transformations// *Library & Information Science Research*. — 2012. — Vol. 34, No.3. — P. 238-246.
36. *Wickramasinghe V., Welivitiyoda P.* Benefits gained from dimensions of social capital: A study of software developers in Sri Lanka// *Information Technology & People*. — 2011. — Vol. 24, No.4. — P. 393-413.
37. *Wu W.-Y., Chang M.-L., Chen C.-W.* Promoting innovation through the accumulation of intellectual capital, social capital, and entrepreneurial capital// *R & D Management*. — 2008. — Vol. 38, No.3. — P. 265-277.
38. *Huvila I., Holmberg K., Ek S., Widen-Wulff G.* Social capital in Second Life// *Online Information Review*. — 2010. — Vol. 34, No.2. — P. 295-316.
39. *Varheim A.* Public libraries: Places creating social capital// *Library Hi Tech*. — 2009. — Vol. 27, No. 3. — P. 372-381.
40. *Aabo S., Audunson R.* Use of library space and the library as place// *Library & Information Science Research*. — 2012. — Vol. 34, No. 2. — P. 138-149.
41. *Varheim A.* Trust and the role of the public library in the integration of refugees: The case of a northern Norwegian city// *Journal of Librarianship and Information Science*. — 2014. — Vol. 46, No. 1. — P. 62-69.
42. *Aabo S., Audunson R., Varheim A.* How do public libraries function as meeting places// *Library & Information Science Research*. — 2010. — Vol. 32, No. 1. — P. 16-26.
43. *Audunson R.A., Varheim A., Aabo S., Holm E.D.* Public libraries, social capital and low intensive meeting places// *Information Research*. — 2007. — Vol. 12, No. 4, paper colis 20. — <http://www.informationr.net/ir/12-4/colis/colis20.html> (Archived by WebCite at <http://www.webcitation.org/6cNOzKjSw>).
44. *Audunson R.A., Essmat S., Aabo A.* Public libraries: A meeting place for immigrant women? *Library & Information Science Research*. — 2011. — Vol. 33, No. 3. — P. 220-227.
45. *Johnson C.A.* Social capital and the search for information: Examining the role of social capital in information seeking behavior in Mongolia// *Journal of the American Society for Information Science and Technology*. — 2007. — Vol. 58, No. 6. — P. 883-894.
46. *Johnson C.A.* Do public libraries contribute to social capital? A preliminary investigation into the relationship // *Library & Information Science Research*. — 2010. — Vol. 32, No.2. — P. 147-155.
47. *Johnson C.A.* How do public libraries create social capital? An analysis of interactions between library staff and patrons// *Library & Information Science Research*. — 2012. — Vol. 34, No.1. — P. 52-62.
48. *Johnson C.A., Griffis M.R.* A place where everybody knows your name? Investigating the relationship between public libraries and social capital// *The Canadian Journal of Information and Library Science*. — 2009. — Vol. 33, No.3/4. — P. 159-191
49. *Johnson C.A., Griffis M.R.* The effect of public library use on the social capital of rural communities// *Journal of Librarianship and Information Science*. — 2014. — Vol. 46, No.3. — P. 179-190.
50. *Griffis M.R., Johnson C.A.* Social capital and inclusion in rural public libraries: A qualitative approach// *Journal of Librarianship and Information Science*. — 2014. — Vol. 46, No.2. — P. 96-109
51. *SK 2006 Community Benchmark Survey*. — Boston, MA: The Saguaro Seminar, Harvard Kennedy School, Harvard University, n.d. — <http://www.hks.harvard.edu/programs/saguaro/measurement/sk-2006-community-benchmark-survey>. (Archived by WebCite at <http://www.webcitation.org/6cNP6Shah>).
52. *Wellman B.* The community question: The intimate networks of East Yorkers// *American Journal of Sociology*. — 1979. — Vol. 84, No. 5. — P. 1201-1231.
53. *Wellman B., Quan-Haase A., Witte J., Hampton K.* Does the Internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment// *American Behavioral Scientist*. — 2001. — Vol. 45, No.3. — P. 437-456.
54. *Ellison N.B., Steinfield C., Lampe C.* The benefits of Facebook “friends:” Social capital and college students' use of online social network sites// *Journal of Computer Mediated Communication*. — 2007 — Vol. 12, No. 4. — P. 1143-1168.
55. *Wellman B., Salaff J., Dimitrova D., Garton L., Garton M., Haythornthwaite C.* Computer networks as social networks: collaborative work, telework, and virtual community// *Annual Review of Sociology*. — 1996. — Vol. 22. — P. 213-238.
56. *Parks M. R., Floyd K.* Making friends in cyberspace// *Journal of Computer-Mediated Communication*. — 1996. — Vol. 1, No. 4. — <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1996.tb00176.x/full>. (Archived by WebCite at <http://www.webcitation.org/6cNPnheDh>)
57. *Ahn J.* Teenagers' experiences with social network sites: Relationships to bridging and bonding social capital// *The Information Society*. — 2012. — Vol. 28, No. 2. — P. 99-109.
58. *Williams D.* On and off the ‘net: Scales for social capital in an online era// *Journal of Computer-Mediated Communication*. — 2006. — Vol. 11, No. 2. — P. 593-628.
59. *Appel L., Dadlani P., Dwyer M., Hampton K., Kitzje V., Matni Z.A., Moore P., Teodoro R.* Testing the validity of social capital measures in the study of information and communication technologies// *Information, Communication & Society*. — 2014. — Vol. 17, No.4. — P. 398-418.
60. *Gaag M. van der, Snijders T.A.B.* The resource generator// *Social Networks*. — 2005. — Vol. 27, No. 1. — P. 1-29.

61. *Loane S.S., D'Alessandro S.* Peer-to-peer value through social capital in an online motor neuron disease community// *Journal of Nonprofit & Public Sector Marketing*. — 2013. — Vol. 25. — P. 164-185.
62. *Huysman M., Wulf V.* IT to support knowledge sharing in communities: towards a social capital analysis// *Journal of Information Technology*. — 2006. — Vol. 21, No. 1. — P. 40-51.
63. *Yang S.-C., Farn C.-K.* Social capital, behavioral control, and tacit knowledge sharing- a multi-informant design//*International Journal of Information Management*. — 2009. — Vol. 29, No. 3. — P. 210-218.
64. *Hau Y.S., Kim B., Lee H., Kim Y-G.* The effects of individual motivations and social capital on employees' tacit and explicit knowledge sharing intentions// *International Journal of Information Management*. — 2013. — Vol. 33, No. 2. — P. 356-366.
65. *Zhang L., Jones M.C.* Can social capital enhance the careers of IT professionals// *Information Resources Management Journal*. — 2009. — Vol. 22, No. 2. — P. 69-82.
66. *Bhandar M.* A framework for knowledge integration and social capital in collaborative projects// *Electronic Journal of Knowledge Management*. — 2010. — Vol. 8, No. 3. — P. 267-280.
67. *Warschauer M.* Social capital and access// *Universal Access in the Information Society*. — 2003. — Vol. 2, No. 4. — P. 315-330.
68. *Williams K.* Informatics moments// *Library Quarterly*. — 2012. — Vol. 82, No. 1. — P. 47-73.
69. *Chen W.* The implications of social capital for the digital divides in America// *The Information Society*. — 2013. — Vol. 29, No. 1. — P. 13-25.
70. *Farooq U., Ganoe C.H., Xiao L., Merkel C.B., Rosson M.B., Carroll J.M.* Supporting community-based learning: Case study of a geographical community organization designing its website// *Behaviour & Information Technology*. — 2007. — Vol. 26, No. 1. — P. 5-21.
71. *Mignone H.* Impact of ICT on social capital in aboriginal communities in Canada// *Journal of Information, Information Technology & Organizations*. — 2009. — Vol. 4. — P. 127-145.
72. *Hui Y., Wenjie Z., Shenglong H.* Social capital, digital inequality, and a "glocal" community informatics project in Tianshu Tibetan Autonomous County, Gansu Province // *Library Trends*. — 2013. — Vol. 62, No.1. — P. 234-260.
73. *Johnson C.A.* Choosing people: The role of social capital in information seeking behaviour// *Information Research*. — 2004. — Vol. 10, No.1, paper 201. — <http://www.informationr.net/ir/10-1/paper201.html> (Archived by WebCite at <http://www.webcitation.org/6cNOtYTmE>)
74. *Woudstra L., van den Hoof B., Schouten A.P.* Dimensions of quality and accessibility: Selection of human information sources from a social capital perspective// *Information Processing & Management*. — 2012. — Vol. 48, No. 1. — P. 618-630.
75. *Case D.O.* Looking for information: A survey of research on information seeking, needs, and behavior. — Amsterdam, The Netherlands: Academic Press, 2002.
76. *Veinot T.C.* Interactive acquisition and sharing: Understanding the dynamics of HIV/AIDS information networks// *Journal of the American Society for Information Science and Technology*. — 2009. — Vol. 60, No. 11. — P. 2313-2332.
77. *Veinot T.C.* A multi-level model of HIV/AIDS information/help network development// *Journal of Documentation*. — 2010. — Vol. 66, No.6. — P. 875-890.
78. *Glaeser E., Laibson D., Scheinkman J., Soutter C.* Measuring trust// *Quarterly Journal of Economics*. — 2000. — Vol. 115, No. 3. — P. 811-846.
79. *Naef M., Schupp J.* Measuring trust: Experiments and surveys in contrast and combination. — Bonn, Germany: Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor). (IZA Discussion Paper No. 4087). — 2009. — <http://ftp.iza.org/dp4087.pdf> (Archived by WebCite at <http://www.webcitation.org/6cNSe5G4i>).

Фокус-фактор: динамичное измерение специализации журналов*

Йеппе НИКОЛАЙСЕН
(Jeppe NICOLAISEN)

Копенгагенский университет,
г. Копенгаген, Дания

Тове Фабер ФРАНДСЕН
(Tove Faber FRANDSEN)

Университетская клиника Оденсе,
г. Оденсе, Дания

Представляется новый библиометрический показатель измерения специализации журналов во времени, названный фокус-фактором. Этот показатель основан на библиографическом сочетании и подсчетах доли повторных ссылок, приведенных в последующие годы. Применимость данного показателя демонстрируется на выборке общенаучных журналов и выборке журналов по медицине. Библиографические списки каждого журнала сравнивались по годам, а доля повторных ссылок подсчитывалась путем деления числа повторных ссылок на общее число ссылок за каждый год. Для обоснования повторных ссылок, вызванных специализацией, были измерены и сопоставлены другие возможные причины (старение, самоцитирования журнала и число ссылок). Результаты показывают, что фокус-фактор способен проводить различие между журналами общего профиля и специальными журналами, а, значит, он эффективно измеряет запланированное явление (т.е. специализацию журналов). Только слабые корреляции были найдены между повторными ссылками журналов и старением, самоцитированиями журналов и числом ссылок. Фокус-фактор успешно измеряет специализацию журнала во времени. Измерения, основанные либо на простом анализе ссылок, либо на библиографическом сочетании, считаются тесно связанными. Измерения, основанные на самоцитировании журнала, только слабо коррелируют с фокус-фактором. Измерения, основанные на анализе цитирования, нуждаются в изучении и сравнении.

ВВЕДЕНИЕ

Журнал *The Philosophical Transactions of the Royal Society of London* («Философские труды Лондонского королевского общества») обычно считается первым научным журналом. Он был опубликован впервые в 1665 г., тремя месяцами позже, чем французский журнал *Journal des Sçavans* («Журнал ученых»). Оба журнала публиковали научную информацию, но *Journal des Sçavans* больше фокусировался на обзоре книг, чем на передаче нового исследования в форме статьи [1]. Эти два журнала, а также последовавшие за ними журналы возникли как публичные записи деятельности и интересов научных обществ. Журналы, не имеющие отношения к научным сообществам, начали появляться на протяжении восемнадцатого и девятнадцатого столетий [1]. В ответ на быстрый

рост результатов новых исследований в течение девятнадцатого века ученые стали специализироваться в подобластях. Ученые уже больше не могли следить за развитием в их собственной дисциплине в целом и вместо этого сконцентрировались на меньших дисциплинарных разделах. Это также привело к росту специализации многих научных журналов [1].

Таким образом, длительное время научный журнал был самым важным средством передачи новых идей и знаний в науке, в большинстве социальных наук, и в меньшей, но увеличивающейся степени в искусстве и гуманитарных науках. Возможно из-за этого статуса самого важного средства научной коммуникации научный журнал как средство передачи результатов исследований завоевал интерес ряда областей (например, истории науки, социологии науки, лингвистики, библиотековедения и информатики и др.). Тем не менее, область, больше всего изучающей научный журнал, несомненно является библиометрия. На ум приходит одно исследование: основополагающее открытие Дерекком Дж. де Солла Прайсом экспоненциального роста науки, которое было основано на исследованиях самого старого научного журнала (*Philosophical Transactions*) и других жур-

* Перевод Nicolaisen J., Frandsen T.F. The focus factor: A dynamic measure of journal specialization.—
<http://www.informationr.net/ir/20-4/paper693.html>

налов [2]. Последовали другие многочисленные библиометрические исследования, сфокусированные на научном журнале (их слишком много, чтобы упоминать здесь). Некоторые из этих исследований стремились установить адекватные измерения на основе цитирований для различных аспектов научного журнала. Среди наиболее важных находится исследование Бартона и Кеблера [3], в котором они разработали измерение старения научной литературы (период полураспада), позднее используемое для измерения старения научных журналов (например, The Journal Citation Reports), а также исследование Гарфилда и Шера [4], раскрывшее суть импакт фактора журнала. Многие альтернативные измерения старения журналов и их влияния разрабатывались с 1960-х гг., но важный аспект научного журнала – журнальная специализация, таким образом, во многом не принимался во внимание или игнорировался этим направлением исследования библиометрического показателя. Это действительно немного странно, так как специалисты по библиометрии давно знали о важном факторе специализации для результата библиометрических исследований. Почти сорок лет назад Генри Смолл указал на основной результат экспериментов, проведенных одним - двумя годами ранее ведущими библиометриками, включая помимо других Юджина Гарфилда, Бельвера С. Гриффита и его самого, и пришел к выводу, что *первоначальной структурной единицей в науке является научная специализация* [5, с. 67]. Общая критика журнального импакт фактора состоит в том, что импакт фактор журнала частично определяется его уровнем специализации (например, [6]). Таким образом, чтобы улучшить нашу интерпретацию библиометрических показателей журналов (например, импакт фактора журнала), нам нужно простое, но эффективное измерение журнальной специализации. Такое, которое могло быть быстро встроено в продукты, например, такие как The Journal Citation Reports.

Николаисен и Франнсен [7] разработали простое, но эффективное измерение для оценки специализации научных журналов. Они представили идею этого нового, основанного на ссылках измерения на конференции COLIS8-conference (г. Копенгаген) в 2013 г. [7] и в качестве краткого сообщения опубликовали в журнале *Journal of the Association for Information Science and Technology* [8]. Протестировав это измерение на большой выборке, мы приводим более подробное исследование этого нового измерения. Специализация приравнивается к сужению фокуса, и поэтому мы выбрали *фокус-фактор* в качестве названия для этого нового, основанного на ссылках журнального измерения.

В следующем разделе представлено детальное описание фокус-фактора, включая основные теоретические предположения, на которых он базируется, родственные измерения и то, как он подсчитывается. Далее показывается применение этого нового показателя на выборке научных журналов и проверяется его обоснованность в качестве измерения специализации журналов.

ТЕМАТИЧЕСКИ СВЯЗАННАЯ ЛИТЕРАТУРА И РОДСТВЕННЫЕ ИЗМЕРЕНИЯ

Новое, основанное на ссылках измерение, которое мы собираемся представить, является измерением специализации журналов. Оно основано на общем определении научных специальностей и специализации, которое можно найти во многих текстах по социологии науки и исследованиях науки (включая библиометрию). Перед

представлением фокус-фактора в деталях кратко обрисовуем это традиционное понимание и определение и затронем родственные библиометрические измерения.

Специальности и специализация

В своей книге «Communicating research» Медоуз [9] помимо других тем обсуждает быстрый рост научных исследований и то, как научное сообщество развивает механизм, помогающий справиться с избытком информационной продукции. Этим механизмом, по мнению Медоуза [9, с. 20], является специализация. Чтобы точно понять, что он подразумевает под специализацией, необходимо подробнее изучить его довод. Медоуз [9, с. 19] просит читателя выслушать жалобу Фарадея 1826 г.:

Любому человеку, желающему посвятить часть своего времени химическому опыту, конечно невозможно прочитать все книги и статьи, опубликованные относительно его цели; их число огромно, а труд, чтобы отобрать несколько экспериментальных и теоретических истин, которые во многих из них приводят в замешательство очень большой долей не представляющих интереса вопросов, фантазий и ошибок, является таким, что большинство людей, пытающихся воспроизвести опыты, быстро убеждается, что надо делать выбор в своем чтении и при этом иногда случайно не пройти мимо того, что на самом деле заслуживает внимания.

Сегодня имеется гораздо больше информации, с которой надо уметь справляться, чем во времена Фарадея. Поэтому можно логично прийти к убеждению, что проблема, которую описал Фарадей, в наше время становится более острой. Однако, по мнению Медоуза [9], это не так. Причина заключается в том, что современные химики больше не пытаются распорядиться тем, что Медоуз [9, с.20] назвал тем же самым *широким размахом их области*, как это делали химики во времена Фарадея. Вместо этого современные химики концентрируются на гораздо более ограниченных темах [9, с. 20]. Ученые становятся более *специализированными* [9, с. 20]. Если исследование расширяется, ученые ограничивают свое внимание его определенными частями [9, с. 20]. Поэтому ученые любой дисциплины обычно интересуются только частью своей области [9, с. 21].

Это определение специальностей отражает идею общественного разделения труда в обществе. Во всех известных обществах производство товаров и услуг делится на разные рабочие задачи таким образом, чтобы никто из членов общества не выполнял всех задач. Наоборот, типы рабочих задач, которые может выполнять индивидuum, часто регулируются правилами, и индивидуумы обязаны выполнять определенные задачи. Адам Смит (1723-1790) был первым, кто сформулировал, как общественное разделение труда ведет к росту производительности. В своей книге *On the wealth of nations*, опубликованной в 1776 г., он даже утверждал, что разделение труда является самой важной причиной экономического роста. Знаменитый пример из его книги иллюстрирует его точку зрения. Пример связан с заводом по производству булавок. Согласно теории Смита, завод по производству булавок, который применяет разделение труда, может производить десятки тысяч булавок в день, тогда как завод по производству булавок, где каждый рабочий пытается выпускать булавки от начала и до конца, выполняя все задачи, связанные с их производством, будет производить очень мало булавок. То, что Медоуз по-видимому имел в виду при описании страте-

гин, адаптированной современными химиками, является таким образом стратегией успеха завода по производству булавок. Подобно рабочим успешного завода по производству булавок, современные химики разделили свои рабочие задачи между собой и, следовательно, работают над различными, но родственными задачами. Сегодня в химии имеется несколько специальностей, включая органическую химию, неорганическую химию, химическое машиностроение и многое другое. То же самое считается верным для всех других научных областей. Социологи, например, обычно работают в рамках одной из специальностей, описанных в книге Смелсера «Handbook of sociology» [10]. Они включают, помимо прочего, социологию образования, социологию религии, социологию науки, медицинскую социологию, социологию средств массовой информации, социологию возраста и социологию отношений полов.

Медоуз [9, с. 44] напоминает, что дисциплины и специальности также могут производиться за счет слияния. Сочетание части биологии с частью химии для появления биохимии является одним из примеров.

Следовательно, то, что характеризует специальность, является феноменом или феноменами, которые изучают специалисты определенной области. Органическая и неорганическая химия, например, являются разными специальностями, потому что их ученые изучают различные феномены. Химики-органики изучают материалы, основанные на углероде, такие как нефть или уголь, тогда как химики-неорганики работают с материалами, не содержащими ни углерода, ни синтетические материалы на основе углерода. Социологи науки изучают научные сообщества, в то время как социологи религии изучают религиозные сообщества. Хотя большая часть членов этих двух групп обучается дисциплине социологии, они принадлежат к разным социологическим специальностям, поскольку изучают различные социологические феномены.

Как отмечалось выше, определение специальностей Медоуза соответствует определению, обычно применяемому в научных исследованиях. Крейн и Смолл [11, с. 198], например, объясняют понятие специальностей, утверждая, что

кластеры родственных научных областей содержат специальности, обладатели которых связаны друг с другом наличием общего интереса к конкретному типу феномена или метода (таким как преступление, семейные узы, популяции и т.д.). Дисциплины в свою очередь состоят из кластеров специальностей.

Смолл и Гриффит [12, с. 17] поддерживают идею, что «наука является мозаикой специальностей, а не единым целым», и отмечают, что специальности служат *строительными блоками* науки. Герин [13], Уитли [14] и Цукерман [15] утверждают, что область проблемы состоит из ряда родственных, но отдаленных проблем, что кластер родственных областей проблемы включает специальность и что научная дисциплина охватывает множество родственных специальностей.

Измерение специализации

Хагстром [16, с. 91-92] утверждает — *«есть основание верить, что ученые продолжают общаться чаще и интенсивнее с другими коллегами в их специальностях, обмениваясь с ними препринтами, цитируя их работы и обмениваясь репринтами»*. Циман [17, с. 190] отмечает, что «научные специальности

часто кажутся закрытыми друг от друга стеной взаимного игнорирования». Эти предположения изучаются и эмпирически подтверждаются библиометриками.

Используя простой анализ ссылок, Эрл и Викери [18] исследовали, в какой степени разнообразие предметных областей описано в литературе их собственной области (предметное самоцитирование) и других предметных областях. Они обнаружили значительные колебания между предметными областями в ходе исследования, которые по-видимому должны соответствовать предположениям относительно специализации. Среди их наблюдений было такое, что значительная зависимость от других предметных областей найдена в естественных областях и областях общей технологической направленности, тогда как математика, по их наблюдениям, мало зависела от литературы других предметных областей.

Авторские самоцитирования также находят отражение в тенденциях специализации. Часто авторские самоцитирования подвергаются критике со стороны анализа цитирования (например, [19-21]). Критики предполагают или даже утверждают (см. [19, с. 636]), что авторские самоцитирования равноценны саморекламе и поэтому должны исключаться из оценок библиометрии. Тем не менее, исследование пятидесяти одного самоцитирующего автора, проведенное Бонци и Снайдером [22], открыло, что по существу нет различий в причинах того, что авторы цитируют свои собственные работы, и того, что они цитируют работы других. Самоцитирования в первую очередь определяют тематически связанную или более раннюю работу, на которую опираются поздние работы. Таким образом, авторские самоцитирования кажутся подчеркивают специальный акцент автора на узкопрофильную научную проблему. Ранние исследования Паркера, Пейсли и Гарретта [23], Медоуза и О'Коннера [24] также задокументировали подобные взаимосвязи между специализацией и авторскими самоцитированиями.

Метод коцитирования Маршаковой [25] и Смолла [26] предоставляет количественную технику группирования или кластеризации цитируемых документов или цитируемых авторов. Измеряя силу коцитирования в достаточной большой выборке единиц (например, документов или авторов), можно выявить кластеры единиц, которые являются высоко коцитируемыми. Ученые информатики, которые заинтересовались этой техникой в 1970-х гг. и позже, повторно обнаружили, что такие кластеры адекватно представляют научные специальности (например, [12, 27-29]).

Библиографическое сочетание ([30]) является родственным методом кластеризации связанных объектов. Документы (или другие единицы анализа) считаются библиографически сочетаемыми, если они обмениваются библиографическими ссылками. Библиометрики стали проявлять интерес к этой технике в течение 1990-х гг., используя ее для определения и отображения кластеров предметно-родственных документов (например, [31-33]). Как показано в исследовании Николаисена и Франнсена [34], библиографическое сочетание обладает другим многообещающим потенциалом в качестве измерения уровня консенсуса и специализации в науке. Используя измененную форму библиографического сочетания (агрегированное библиографическое сочетание), они смогли измерить уровень согласованности в двух различных дисциплинах в определенное время.

Фокус-фактор

Специализация является процессом. Уровень специализации внутри дисциплины растет или падает во времени. Чтобы измерить это с помощью библиометрических методов, таких как самоцитирования, анализ коцитирования или библиографического сочетания, нужно включить в него временное измерение. Фокус-фактор создается с этой подразумеваемой целью. Используя научный журнал в качестве единицы выборки, он (фокус-фактор) измеряет уровень специализации путем подсчета перекрытий в библиографических ссылках год за годом. Например: журнал собирает 1536 ссылок в нулевом году и 1622 ссылки в первом году, из которых 219 ссылок находятся в библиографических списках этого журнала за каждый год. Итак, 219 из 1622 ссылок в первом году равноценны ссылкам, найденным в этом же журнале за предыдущий год. Это равняется 13,5% и принимается нами в качестве показателя уровня специализации в отдельном журнале за первый год. Уровень специализации во втором году подсчитывается путем сравнения перекрытия в библиографических ссылках, используемых одним и тем же журналом в первом и во втором годах и т.д.

Данный метод был проверен Николаисеном и Франсененом [7,8] на выборке ядерных журналов в области библиотековедения и информатики. Результаты показали, что фокус-фактор хорошо различает общие и специализированные журналы и, более того, эффективно измеряет уровень специализации среди выбранных журналов.

МЕТОД

Для изучения применимости фокус-фактора в более широком разнообразии предметов и журналов мы проверили данное измерение на выборке общенаучных журналов, общемедицинских журналов и узкоспециализированных медицинских журналов (см. таблицу). Общенаучные журналы включают такие журналы, как *Science*, *Nature* и *PNAS* (см., например, [35]), из которых два первых были выбраны для этого исследования. Общемедицинские журналы, выбранные для нашего исследования, считаются среди всех журналов самыми пре-

стижными медицинскими журналами (см., например, [36]), а также известны как входящие в первую пятерку журналов (например, [37]). Узкоспециализированные медицинские журналы выбраны в качестве примеров из широкого ряда доступных специализированных журналов по совету двух специалистов по информации в сфере медицины (MD и MSC).

Чтобы определить долю повторных ссылок, ссылки определенного года из каждого включенного журнала сравнивались со ссылками в этом же журнале за предыдущий год. Повторная ссылка определяется как 100% соответствие между цитируемой ссылкой в первый год и цитируемой ссылкой за предыдущий год. Это значит, что ошибки орфографии, печатные ошибки, варианты орфографии и подобные несоответствия служат потенциальными источниками противоречий, но поскольку, как ожидается, они должны равномерно распределяться по всему массиву данных, то противоречие маловероятно. Регистрируемыми данными являются название журнала, год издания, цитируемые ссылки в журнале и число примеров для каждой ссылки. Некоторые ссылки появляются более одного раза и поэтому число повторных ссылок зависит от общего числа примеров, а не только от числа единичных ссылок. Информация о журнале, годе издания и цитируемых ссылках в журнале была собрана с использованием базы данных Web of Science. Информация относительно числа примеров для каждой ссылки собиралась при помощи программного обеспечения, специально разработанного для этой цели. Доля повторных ссылок в журнале *j* в году *y* подсчитывается следующим образом:

Доля повторных ссылок = число повторных ссылок (*j,y*)/общее число ссылок (*j,y*).

Ниже дается пример того, как подсчитывается доля повторных ссылок:

в 2011 г. *Nature* содержал 32069 ссылок, из которых 4971 были повторными, приводя к доле повторных ссылок равной $4971/32069=0,155$.

В общем это исследование проанализировало 4788579 ссылок в 15 журналах за 1991-2012 гг. и подсчитало долю повторных ссылок. Включались только статьи, краткие сообщения, обзоры и письма. Письма были включены по рекомендации Кристенсена и др. [38].

Таблица

Список включенных журналов

Общенаучные журналы
Science
Nature
Журналы по общей медицине
British Medical Journal
The Journal of the American Medical Association
Annals of Internal Medicine
Lancet
New England Journal of Medicine
Узкоспециализированные журналы по медицине
Ophthalmology
Archives of Ophthalmology
American Journal of Ophthalmology
British Journal of Ophthalmology
Experimental Eye Research
Investigative Ophthalmology
Journal of Clinical Oncology
JNCI: Journal of the National Cancer Institute

РЕЗУЛЬТАТЫ

Журналы, включенные в анализ, являются в различной степени специализированными. Доля повторных ссылок колеблется от 0,05 до 0,2 и выше, т.е. почти 20% ссылок в любом выбранном году появляются в этом специализированном журнале в предыдущем году. Рис.1 и 2 иллюстрируют развитие уровней специализации за 1991-2012 гг. В *Приложении* приводятся конкретные подсчеты.

Рис. 1 показывает результаты анализов общенаучных журналов и журналов по общей медицине.

Один журнал выделяется на этом рисунке, так как он характеризуется большей степенью специализации, особенно в последние пять-шесть лет. *Nature* кажется более высокоспециализированным, чем другие журналы, изображенные на рис. 1. Эта тенденция по-видимому снижается на протяжении первой декады анализа, но усиливается на протяжении последней. Другой общенаучный журнал, *Science*, начинает свой путь как высокоспециализированный, но движется в сторону меньшего уровня специализации на протяжении всего периода анализа.

Рис. 2 представляет обзор результатов анализа специализированных журналов по медицине. Для сравнения результатов шкалы горизонтальных осей рис. 1 и 2 являются одинаковыми.

Специализированные журналы по медицине проявляют большую вариацию с долей повторных ссылок,

изменяющейся от 0,10 до 0,37. Некоторые журналы являются специализированными на уровне, больше напоминающем общенаучные журналы, тогда как для других журналов 30% ссылок в некоторых годах появляются в этих отдельных журналах за предыдущий год.

Николаисен и Франнсен [7] анализировали, можно ли объяснить уровни повторных ссылок старением. Они проверяли данную гипотезу, изучая возраст распределения ссылок в журнале с помощью измерения возраста периода полураспада или медианного возраста цитирования. Был применен метод дискретного анализа, так как годы издания обрабатывались как дискретные единицы, а не как континуум дат в условиях интервалов. Корреляция была положительной, т.е. журналы, включающие относительно большую долю более старых ссылок, характеризуются большим уровнем специализации – все другие единицы уравнивались. Журналы с относительно новыми ссылками имеют меньше повторных ссылок просто потому, что наблюдается больше ссылок в этих журналах, которые не могли цитироваться годом ранее. Эта гипотеза также проверяется на данных, приводящих к одинаковому результату. Рис. 3 представляет иллюстрацию корреляции и, аналогично предыдущему анализу, R^2 указывает на то, что медианный возраст цитирования сам по себе не объясняет различные уровни повторных ссылок. Конкретные данные приведены в *Приложении*.

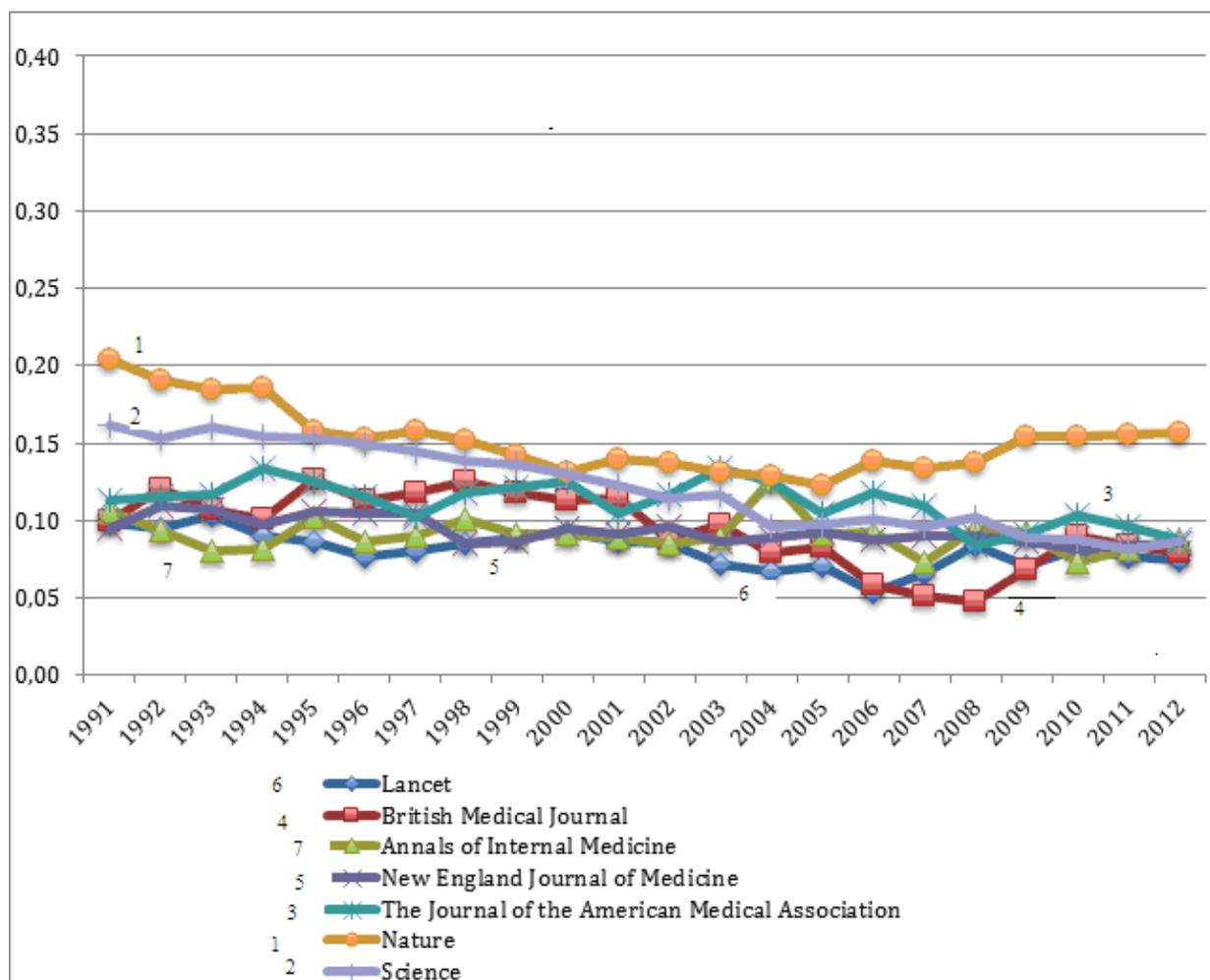


Рис.1. Уровень специализации (общенаучные журналы и журналы по общей медицине)

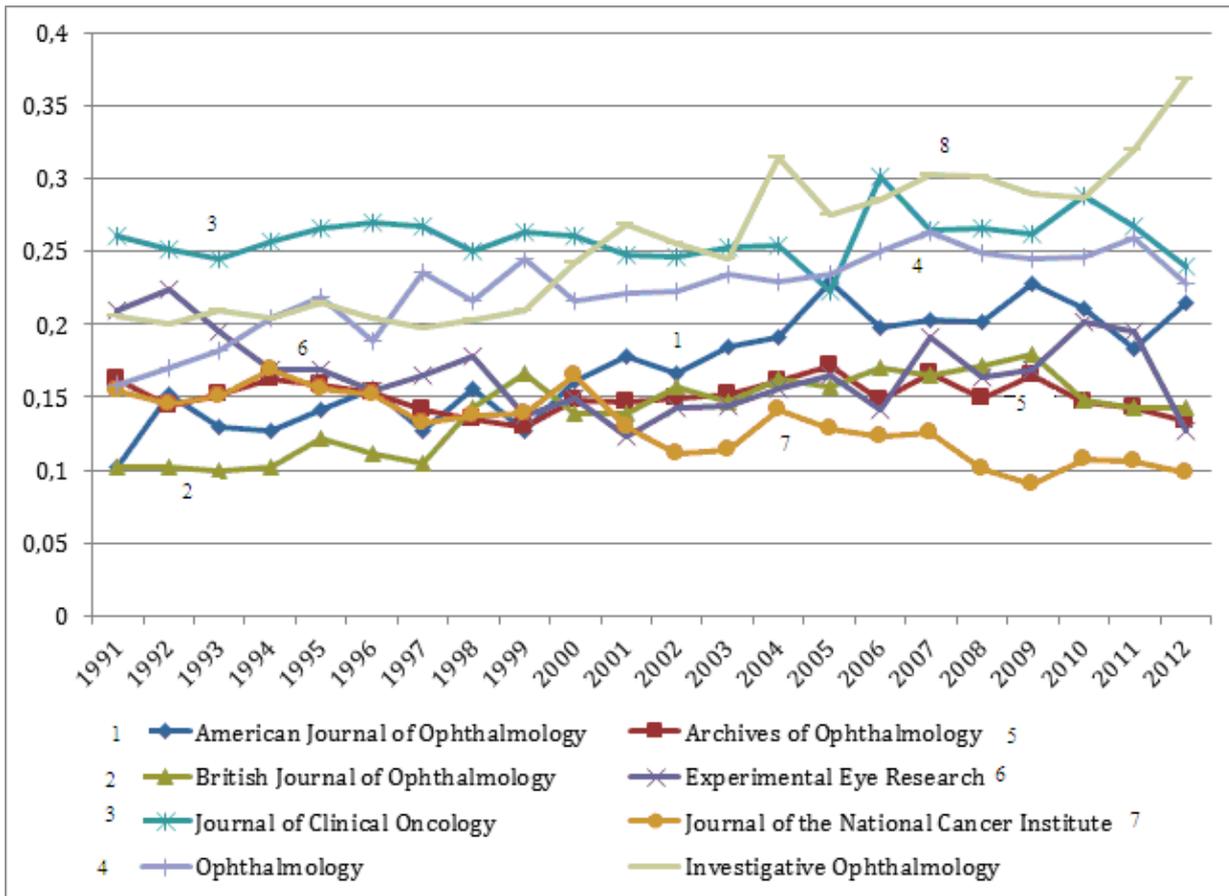


Рис. 2. Уровень специализации (специализированные журналы по медицине)

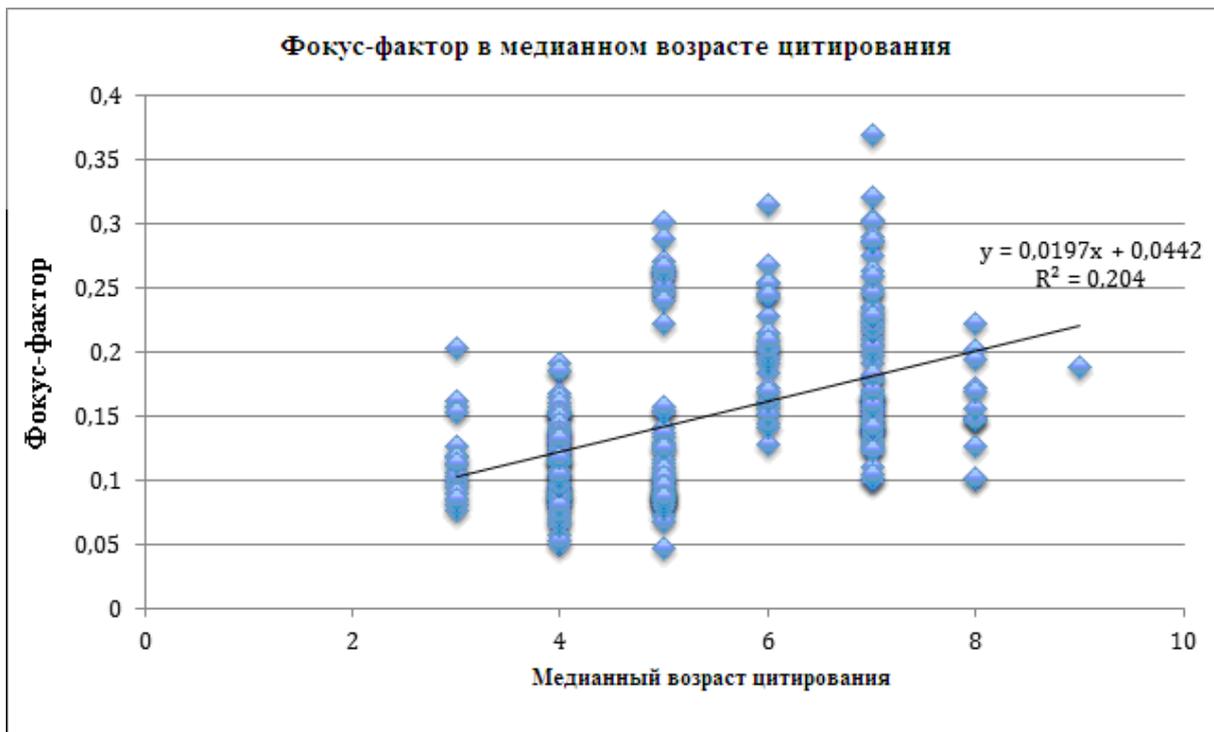


Рис. 3. Медианный возраст цитирования и доля повторных ссылок

Обращаясь к другому общепризнанному используемому измерению специализации, которое, вероятно, могло бы объяснить различия в уровнях повторных ссылок, мы проанализируем корреляцию между долей повторных ссылок и самоцитирований журналов. Доля повторных ссылок измеряется для каждого журнала за целый период времени и соотносится с долей повторных ссылок. Рис. 4 отражает некоторую корреляцию, хотя определенно не самую сильную. Подсчеты данных см. в *Приложении*.

Значение R^2 , равное 0,19, подтверждает, что самоцитирования журналов и доля повторных ссылок не должны считаться одинаковыми измерениями.

Наконец, мы изучаем, вызваны ли различия в уровнях повторных ссылок различиями в числе ссылок. Некоторые утверждают, что более значимые журналы имеют больше ссылок, которые могут повторно цитироваться. Однако это измерение не является абсолютным, а значит, более важные журналы не должны проявлять более высокие уровни повторных ссылок. Доля самоцитирований измеряется для каждого журнала за целый период времени и соотносится с общим числом ссылок, полученных в течение этого года.

Рис. 5 отражает очень слабую корреляцию между числом ссылок и долей повторных ссылок. Данные о подсчетах см. в *Приложении*.

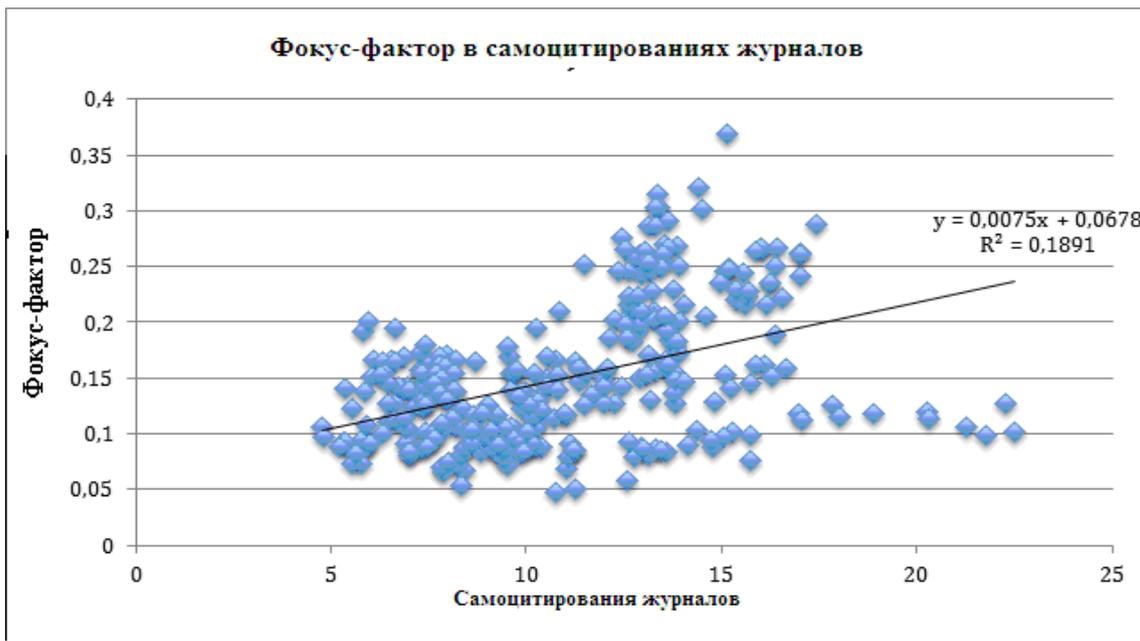


Рис. 4. Самоцитирования журналов и доля повторных ссылок

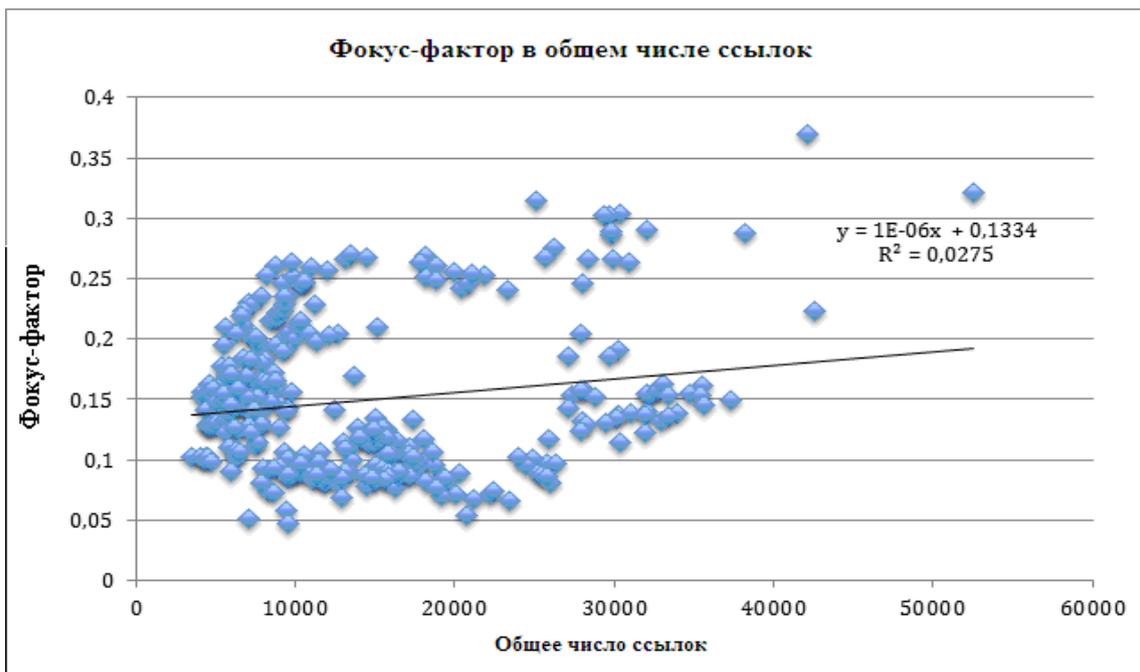


Рис. 5. Число ссылок и доля повторных ссылок

ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

Научные журналы обслуживают разные цели. Среди самых важных из них – доверие, распространение и архивирование научных результатов. Хотя научные журналы, можно сказать, обмениваются важными характеристиками, способ, которым они обслуживают свои цели, средство, с помощью которого они стремятся удовлетворять их, а также успех в их обслуживании являются в высшей степени разнообразными. Итак, поиск развития единственного и уникального измерения научных журналов невозможен. Как отметил Руссо [39], качество журнала является многофасетным понятием, питающим целую плеяду показателей. Фокус-фактор является новым вкладом в эту плеяду показателей.

Измерение уровня специализации демонстрирует новизну в исследовании библиометрического показателя. Мы убеждены, что уровень специализации – важный аспект научных журналов. Кроме того, подобно другим показателям, фокус-фактор измеряет только один аспект научных журналов и становится интересным тогда, когда другие аспекты также учитываются. Более того, как и в случае других показателей, единичное показание не может быть принято в качестве определенного доказательства. Это измерение должно применяться во времени, приводя к нескольким показаниям, которые необходимо сравнивать с оценками других показателей журналов. Только с помощью такого подхода мы получим адекватную картину научного журнала.

При измерении общего уровня специализации в двух группах журналов фокус-фактор вполне способен сделать различие между группами общих и специализированных журналов. Общонаучные журналы, представленные на рис. 1, показывают цифры в диапазоне от 0,05 до почти 0,2. Специализированные журналы, представленные на рис. 2, дают цифры в диапазоне от 0,1 до 0,35. Однако мы нашли перекрывающее показание в двух группах журналов (между 0,1 и 0,2). Особо концентрируясь на *Nature*, мы обнаружили, что он по всей видимости является в некотором роде более специализированным журналом. Однако обычно *Nature* считается общенаучным журналом. Логично, что либо ошибается фокус-фактор, либо оценка *Nature* как общенаучного журнала сама по себе неверна. Мы убеждены в последнем. При применении фокус-фактора к выборке журналов по библиотекведению и информатике Николаисен и Франнсен [7] нашли похожее отклонение: *Journal of the American Association for Information Science and Technology* (JASIST). При измерении уровня специализации с помощью фокус-фактора они обнаружили, что JASIST, как правило, имеет более высокие оценки, чем большинство специализированных журналов в области. JASIST является журналом, который стремится охватить область целиком и поэтому обычно считается общим журналом. При более глубоком взгляде на эту прозрачную аномалию Николаисен и Франнсен [8] обнаружили, что высокие оценки JASIST были в основном вызваны большим составом опубликованных в нем библиометрических статей. Итак, они обнаружили, что дело не в неудачах фокус-фактора, а в том, что JASIST со временем переносит свое внимание больше в сторону библиометрии, становясь таким образом постепенно все более специализированным. То же самое, вероятно, происходит и с *Nature*. Более подробное исследование возможно откроет пару предпочтительных тем этого журнала (например, биология клетки, ядерная физика, астрофизика или даже антропология), привнося спе-

циализированные статьи с более высокими степенями повторных ссылок. Важным следствием этих наблюдений является то, что бинарное понятие общих и специализированных журналов возможно является слишком ограниченным. В реальности существует более богатый выбор.

Раньше специализация измерялась с помощью других критериев:

1. Простой анализ ссылок [18]
2. Авторские самоцитирования [23, 24]
3. Коцитирование [12, 27-29]
4. Библиографическое сочетание [31-34].

В некоторой степени фокус-фактор может считаться дальнейшим развитием пп. 1-4. Вместо концентрации на предметных областях и степени, с которой они полагаются на собственную литературу (как делали Эрл и Викери [18]), фокус-фактор акцентирует внимание на научных журналах и степени, с которой они опираются на собственную литературу (определенную как литературу, используемую и цитируемую годом раньше). Ясно, что такое отношение равноценно библиографическому сочетанию. Итак, наблюдение, что фокус-фактор является адекватным измерением научной специализации, было ожидаемым. Аналогично, измерение уровня специализации путем самоцитирования журналов так же, как и ожидалось, хорошо справилось со своей задачей. Тем не менее, хотя мы можем задокументировать некоторую корреляцию между повторными ссылками журналов и долей самоцитирований журналов, эти измерения не должны считаться одинаковыми.

Наконец, стоит отметить, что система научной коммуникации постоянно меняется. Все время появляются новые средства коммуникации (например, мегажурналы типа PLOS ONE) и читателей обеспечивают новыми инструментами, чтобы быть на современном уровне разработок в их областях интереса (например, RSS fields, searching Google Scholar и т.д.), делая контекст журнала менее заметным (например, [40]). Показатели журналов, такие как фокус-фактор, безусловно являются только релевантными показателями, поскольку научный журнал остается предпочтительным средством научной коммуникации.

Благодарность. Авторы выражают признательность Дэвиду Хаммеру и Анне Поулсен за их профессиональную помощь в сборе данных, а также двум анонимным рецензентам за их важные предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Meadows A.J.* Communication in science. — London: Butterworth, 1974.
2. *de Solla Price D.J.* Little science, big science. — New York, NY: Columbia University Press, 1963.
3. *Burton R.E., Kebler R.W.* The half-life of some scientific and technical literatures// American Documentation. — 1960. — Vol. 11, No.1. — P. 18-22.
4. *Garfield E., Sher I.H.* New factors in the evaluation of scientific literature through citation indexing// American Documentation. — 1963.—Vol. 14, No.3 —P. 195-201.
5. *Small H.G.* Structural dynamics of scientific literature// International Classification. —1976.— Vol. 3, No.2. — P. 67-74.
6. *Seglen P.O.* Why the impact factor of journals should not be used for evaluating research// British Medical Journal. — 1997.— Vol. 314, No.7079. — P. 498-502.

7. Nicolaisen J., Frandsen T.F. Core journals in library and information science: Measuring the level of specialisation over time// Information Research. —2013.—Vol. 18, No. 3, paper S05.— <http://InformationR.net/ir/18-3/colis/paperS05.html> (Archived by WebCite® at <http://www.webcitation.org/6ahaxXJzO>)
8. Nicolaisen J., Frandsen T.F. Bibliometric evolution: Is the Journal of the Association for Information Science and Technology transforming into a specialty journal?// Journal of the Association for Information Science and Technology.—2015.—Vol. 66, No. 5.—P. 1082-1085.
9. Meadows A.J. Communicating research.— San Diego, CA: Academic Press, 1998.
10. Smelser N.J. (Ed.). Handbook of sociology. — Newbury Park, CA: Sage Publications, 1988.
11. Crane D., Small H. American sociology since the seventies: The emerging identity crisis in the discipline/ T.C. Halliday & M. Janowitz (Eds.). Sociology and its publics: The forms and fates of disciplinary organization (pp. 197-234).— Chicago, IL: University of Chicago Press, 1992.
12. Small H., Griffith B.C. The structure of scientific literatures 1: Identifying and graphing specialties// Science Studies. — 1974.— Vol. 4, No.1.— P. 17-40.
13. Gieryn T.F. Problem retention and problem change in science // Sociological Inquiry.— 1978.— Vol. 48, No.3-4.— P. 96-115.
14. Whitley R. Cognitive and social institutionalisation of scientific specialties and research areas/ R. Whitley (Ed.), Social processes of scientific development (pp. 69-95). — London: Routledge and Kegan Paul, 1974.
15. Zuckerman H. Theory choice and problem choice in science// Sociological Inquiry.—1978.— Vol. 48, No.1.— P. 65-95.
16. Hagstrom W.O. Factors related to the use of different modes of publishing research in four scientific fields / C.E. Nelsen & D.K. Pollock (Eds.), Communication among scientists and engineers (pp. 85-124).— Lexington, MA: Lexington Books, 1970.
17. Ziman J.M. Real science: What it is, and what it means.— Cambridge: Cambridge University Press, 2000.
18. Earle P., Vickery B. Subject relations in science/technology literature// Aslib Proceedings. — 1969. — Vol. 21, No. 6. — P. 237-243.
19. Seglen P.O. The skewness of science// Journal of the American Society for Information Science. — 1992. — Vol. 43, No. 9. — P. 628-638.
20. MacRoberts M.H., MacRoberts B.R. Problems of citation analysis: A critical review// Journal of the American Society for Information Science. — 1989. — Vol. 40, No. 5. — P. 342-349.
21. MacRoberts M.H., MacRoberts B.R. Problems of citation analysis// Scientometrics.—1996.—Vol. 36, No. 3. — P. 435-444.
22. Bonzji S., & Snyder H.W. Motivations for citation: A comparison of self citation and citations to others // Scientometrics. — 1991. — Vol. 21. — P. 245-254.
23. Parker E.B., Paisley W.J., Garrett R. Bibliographic citations as unobtrusive measures of scientific communication. — San Francisco, CA: Stanford University, 1967.
24. Meadows A.J., O'Connor J.G. Bibliographical statistics as a guide to growth points in science// Social Studies of Science. — 1971. — Vol. 1, No. 1. — P. 95-99.
25. Marshakova I.V. A system of document connection based on references//Scientific and Technical Information Serial of VINITI. — 1973. — Vol. 6, No.2. — P. 3-8.
26. Small H. Co-citation in the scientific literature: A new measurement of the relationship between two documents// Journal of the American Society of Information Science. — 1973. — Vol. 24, No.4. — P. 265-269.
27. White H.D., Griffith B.C. Author cocitation: A literature measure of intellectual structure// Journal of the American Society for Information Science. — 1981. — Vol. 32, No. 3. — P. 163-171.
28. White H.D., McCain K.W. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995 // Journal of the American Society for Information Science. — 1998. — Vol. 49, No. 4. — P. 327-355.
29. Zhao D.Z., Strotmann A. The knowledge base and research front of information science 2006-2010: An author cocitation and bibliographic coupling analysis// Journal of the Association for Information Science and Technology. — 2014. — Vol. 65, No. 5. — P. 995-1006.
30. Kessler M.M. Bibliographic coupling between scientific papers// American Documentation. — 1963. — Vol. 14, No. 1. — P. 10-25.
31. Glänzel W., Czernow H.J. A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level// Scientometrics. — 1996. — Vol. 37, No. 2. — P. 195-221.
32. Jarneving B. Bibliographic coupling and its application to research-front and other core documents// Journal of Informetrics. — 2007. — Vol. 1, No.4. — P. 287-307.
33. Ahlgren P., Jarneving B. Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping// Scientometrics. — 2008. — Vol. 76, No. 2. — P. 273-290.
34. Nicolaisen J., Frandsen T.F. Consensus formation in science modeled by aggregated bibliographic coupling // Journal of Informetrics. — 2012. — Vol. 6, No. 2. — P. 276-284.
35. Fanelli D. "Positive" results increase down the hierarchy of the sciences// PLOS One. — 2010. — Vol. 5, No. 4. Retrieved from [http://journals.plos.org/plosone/article?id=10,1371/journal.pone.0010068](http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010068) (Archived by WebCite® at <http://www.webcitation.org/6bBqIKtHl>)
36. Choi Y.M., Nakatomi D., Wu J.J. Citation classics and top-cited authors of psoriasis in five high-impact general medical journals, 1970-2012// Dermatology Online Journal. — 2014. — Vol. 20, No. 5. Retrieved from <http://escholarship.org/uc/item/69n5m3v6#page-1> (Archived by WebCite® at <http://www.webcitation.org/6bBqgz8lr>)
37. Wager E. Getting research published: An A-Z of publication strategy. — Oxford: Radcliffe, 2005.
38. Christensen F.H., Ingversen P., Wormell I. Online determination of the journal impact factor and its international properties// Scientometrics. — 1997. — Vol. 40, No.3. — P. 529-540.
39. Rousseau R. Journal evaluation: Technical and practical issues// Library Trends. — 2002. — Vol. 50, No. 3. — P. 418-439.
40. Lozano G.A., Larivière V., Gingras Y. The weakening relationship between the impact factor and papers' citations in the digital age// Journal of the American Society for Information Science and Technology. — 2012. — Vol. 63, No.11. — P. 2140-2145.

American Journal of Ophthalmology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1991	4361	667	443	0,102	7
1992	4122	622	626	0,152	6
1993	4511	669	582	0,129	6
1994	4482	619	572	0,128	7
1995	4386	545	621	0,142	6
1996	4186	558	653	0,156	6
1997	4574	562	581	0,127	7
1998	4480	613	700	0,156	7
1999	4801	575	610	0,127	7
2000	5268	717	849	0,161	6
2001	5380	744	957	0,178	7
2002	5592	761	929	0,166	7
2003	6748	924	1244	0,184	6
2004	7921	1019	1518	0,192	6
2005	7062	973	1621	0,230	7
2006	7593	980	1503	0,198	7
2007	7041	883	1427	0,203	6
2008	7551	925	1525	0,202	6
2009	7276	963	1660	0,228	6
2010	6697	852	1412	0,211	7
2011	8070	1028	1478	0,183	7
2012	8477	1069	1822	0,215	6

Annals of Internal Medicine

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1991	9382	768	991	0,106	4
1992	10945	839	1031	0,094	5
1993	11877	833	960	0,081	5
1994	11735	830	964	0,082	5
1995	10396	733	1062	0,102	5
1996	9907	725	861	0,087	5
1997	10055	745	902	0,090	5
1998	9518	707	961	0,101	5
1999	9189	644	838	0,091	5
2000	8310	575	755	0,091	5
2001	10333	612	914	0,088	5
2002	9533	553	807	0,085	5
2003	10742	584	958	0,089	5
2004	9036	601	1142	0,126	5
2005	9998	586	918	0,092	5
2006	7984	426	737	0,092	5
2007	8310	478	612	0,074	5
2008	8952	516	845	0,094	5
2009	8625	464	794	0,092	4
2010	8634	480	628	0,073	5
2011	7892	444	641	0,081	5
2012	9579	502	842	0,088	5

Archives of Ophthalmology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	5304	836	766	0,144	6
1993	4911	799	743	0,151	6
1994	4581	728	743	0,162	6
1995	4814	802	765	0,159	7
1996	5281	730	810	0,153	7
1997	5265	803	743	0,141	7
1998	5226	719	708	0,135	7
1999	5025	663	654	0,130	7
2000	5074	702	748	0,147	8
2001	5688	798	837	0,147	8
2002	5561	718	832	0,150	7
2003	6218	813	944	0,152	7
2004	6498	885	1046	0,161	7
2005	5716	794	983	0,172	8
2006	6024	721	895	0,149	8
2007	6154	662	1023	0,166	7
2008	6806	715	1015	0,149	7
2009	6736	759	1111	0,165	7
2010	6467	733	950	0,147	8
2011	5914	709	844	0,143	7
2012	5662	585	755	0,133	7

British Journal of Ophthalmology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	4084	346	415	0,102	8
1993	4315	330	431	0,100	7
1994	4546	348	464	0,102	7
1995	5687	420	693	0,122	7
1996	5852	379	647	0,111	7
1997	6255	432	653	0,104	7
1998	7168	498	1020	0,142	7
1999	6613	432	1098	0,166	7
2000	7283	524	1007	0,138	7
2001	7861	540	1094	0,139	7
2002	7696	572	1208	0,157	8
2003	8968	680	1311	0,146	7
2004	8733	655	1419	0,162	7
2005	8760	636	1382	0,158	7
2006	8277	658	1407	0,170	6
2007	8244	675	1365	0,166	7
2008	8646	630	1485	0,172	6
2009	7873	583	1415	0,180	7
2010	8394	621	1244	0,148	7
2011	9594	729	1367	0,142	7
2012	7133	540	1015	0,142	7

British Medical Journal

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	14323	2900	1718	0,120	3
1993	11621	2473	1238	0,107	3
1994	15556	3501	1572	0,101	3
1995	13888	3094	1759	0,127	3
1996	14636	2972	1661	0,113	4
1997	15562	2942	1828	0,117	3
1998	15719	2801	1973	0,126	4
1999	14533	2465	1722	0,118	4

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
2000	13650	2331	1536	0,113	3
2001	12999	2344	1487	0,114	4
2002	13440	1985	1184	0,088	4
2003	11495	1729	1121	0,098	4
2004	12710	1624	1001	0,079	4
2005	11117	1511	922	0,083	4
2006	9424	1186	549	0,058	4
2007	7127	803	364	0,051	4
2008	9546	1026	450	0,047	5
2009	12912	1426	885	0,069	5
2010	13139	1460	1189	0,090	5
2011	11880	1339	1004	0,085	5
2012	14456	1600	1136	0,079	5

Experimental Eye Research

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	6773	854	1514	0,224	7
1993	5561	571	1083	0,195	6
1994	5741	605	972	0,169	6
1995	5721	548	969	0,169	7
1996	5650	542	873	0,155	7
1997	7565	658	1246	0,165	7
1998	5896	561	1048	0,178	7
1999	5844	478	805	0,138	7
2000	5569	424	833	0,150	7
2001	6392	457	790	0,124	7
2002	6188	417	880	0,142	7
2003	6660	512	956	0,144	7
2004	9746	751	1518	0,156	7
2005	8022	490	1328	0,166	7
2006	12497	669	1762	0,141	7
2007	9490	551	1818	0,192	7
2008	8245	523	1353	0,164	7
2009	13675	941	2319	0,170	8
2010	9344	555	1881	0,201	8
2011	8770	584	1712	0,195	8
2012	7858	509	998	0,127	8

Investigative Ophthalmology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	10035	1397	2011	0,200	7
1993	10181	1353	2134	0,210	6
1994	12631	1674	2582	0,204	6
1995	10389	1460	2234	0,215	6
1996	10885	1591	2227	0,205	6
1997	11367	1426	2251	0,198	6
1998	12094	1605	2450	0,203	6
1999	15136	1959	3171	0,210	6
2000	20404	2647	4947	0,242	6
2001	18220	2526	4888	0,268	6
2002	20017	2705	5106	0,255	6
2003	28026	3693	6887	0,246	6
2004	25145	3359	7910	0,315	6
2005	26284	3276	7250	0,276	7
2006	29810	3914	8527	0,286	7
2007	30342	4062	9202	0,303	7

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
2008	29703	3958	8974	0,302	7
2009	32114	4383	9320	0,290	7
2010	38252	5091	10979	0,287	7
2011	52556	7582	16855	0,321	7
2012	42132	6379	15560	0,369	7

Journal of Clinical Oncology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	8251	947	2079	0,252	5
1993	10615	1343	2606	0,246	5
1994	11970	1528	3080	0,257	5
1995	13082	1639	3479	0,266	5
1996	13508	1831	3653	0,270	5
1997	14480	1973	3868	0,267	5
1998	18144	2527	4549	0,251	5
1999	17824	2326	4698	0,264	5
2000	18828	2538	4914	0,261	5
2001	18888	2544	4685	0,248	5
2002	20896	2585	5149	0,246	5
2003	21816	2850	5509	0,253	5
2004	21055	2762	5344	0,254	6
2005	42562	5466	9488	0,223	5
2006	29376	4265	8870	0,302	5
2007	29974	4815	7957	0,265	5
2008	28331	4533	7554	0,267	5
2009	30885	5252	8119	0,263	5
2010	29775	5189	8590	0,288	5
2011	25727	4226	6870	0,267	5
2012	23360	3977	5617	0,240	5

Journal of the National Cancer Institute

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	5976	386	866	0,145	4
1993	7679	466	1161	0,151	4
1994	6945	541	1172	0,169	4
1995	6483	476	1007	0,155	4
1996	6986	438	1065	0,152	4
1997	7781	617	1026	0,132	4
1998	7451	436	1021	0,137	4
1999	9424	661	1313	0,139	4
2000	8730	578	1439	0,165	4
2001	7888	580	1020	0,129	5
2002	7558	502	842	0,111	5
2003	7678	529	877	0,114	5
2004	7459	555	1054	0,141	5
2005	7069	527	903	0,128	5
2006	7171	396	879	0,123	5
2007	6239	464	786	0,126	5
2008	6346	400	636	0,100	5
2009	5995	359	542	0,090	5
2010	6287	372	672	0,107	5
2011	6520	310	692	0,106	5
2012	4741	228	462	0,097	5

Lancet

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	17735	2614	1673	0,094	3
1993	16980	2437	1750	0,103	3
1994	16924	2392	1525	0,090	3
1995	15848	2110	1365	0,086	3
1996	16238	2558	1238	0,076	3
1997	19403	2549	1565	0,081	3
1998	19294	2593	1646	0,085	3
1999	20352	2639	1797	0,088	4
2000	17515	2210	1612	0,092	4
2001	17046	2212	1482	0,087	4
2002	19134	2149	1623	0,085	4
2003	22174	2111	1583	0,071	4
2004	21153	1785	1422	0,067	4
2005	20026	1651	1422	0,071	4
2006	20725	1730	1110	0,054	4
2007	23483	1852	1537	0,065	4
2008	19404	1719	1630	0,084	4
2009	19173	1497	1344	0,070	4
2010	18130	1689	1476	0,081	4
2011	18866	1783	1449	0,077	4
2012	22419	1793	1661	0,074	4

Nature

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	30289	4110	5783	0,191	4
1993	29756	3603	5510	0,185	4
1994	27149	3414	5046	0,186	4
1995	28094	3379	4429	0,158	3
1996	27319	3148	4198	0,154	4
1997	27920	3168	4408	0,158	4
1998	28768	3112	4382	0,152	4
1999	27159	2855	3857	0,142	4
2000	32966	3321	4328	0,131	4
2001	32088	3089	4473	0,139	4
2002	30279	3019	4160	0,137	4
2003	28049	2838	3686	0,131	4
2004	28292	2917	3652	0,129	4
2005	31893	3155	3916	0,123	4
2006	31105	3355	4321	0,139	4
2007	30179	2990	4047	0,134	5
2008	31998	3159	4391	0,137	5
2009	32416	3309	5001	0,154	5
2010	32133	3124	4940	0,154	5
2011	32069	3096	4971	0,155	5
2012	32787	3194	5156	0,157	5

New England Journal of Medicine

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	17138	1767	1881	0,110	4
1993	18572	1785	1977	0,106	4
1994	17487	1735	1694	0,097	4
1995	17118	1625	1801	0,105	4
1996	16245	1639	1699	0,105	4
1997	15770	1617	1654	0,105	4

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1998	15719	1478	1336	0,085	5
1999	16487	1565	1417	0,086	5
2000	15488	1502	1476	0,095	4
2001	16184	1511	1467	0,091	5
2002	15026	1506	1437	0,096	5
2003	17181	1734	1481	0,086	4
2004	16318	1674	1444	0,088	4
2005	15508	1534	1435	0,093	4
2006	14332	1488	1257	0,088	4
2007	15873	1616	1434	0,090	4
2008	16550	1530	1493	0,090	4
2009	15823	1438	1359	0,086	4
2010	14992	1473	1227	0,082	4
2011	15602	1550	1307	0,084	5
012	14801	1663	1258	0,085	5

Ophthalmology

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	5950	782	1015	0,171	7
1993	7178	994	1311	0,183	7
1994	6331	858	1297	0,205	7
1995	6683	1026	1467	0,220	7
1996	9179	1503	1733	0,189	9
1997	7861	1277	1852	0,236	7
1998	8863	1432	1914	0,216	7
1999	9234	1395	2267	0,246	7
2000	8719	1361	1883	0,216	7
2001	8851	1467	1966	0,222	7
2002	9210	1443	2052	0,223	8
2003	9534	1550	2241	0,235	7
2004	9312	1433	2131	0,229	7
2005	9367	1402	2197	0,235	7
2006	9890	1621	2481	0,251	7
2007	9829	1562	2592	0,264	7
2008	10416	1584	2594	0,249	7
2009	10382	1617	2543	0,245	6
2010	10545	1602	2602	0,247	7
2011	10970	1866	2852	0,260	7
2012	11186	1756	2551	0,228	7

Science

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	33424	2455	5135	0,154	3
1993	35525	2822	5709	0,161	4
1994	34692	2670	5343	0,154	4
1995	35433	2880	5439	0,154	4
1996	37300	2921	5587	0,150	4
1997	35692	2731	5171	0,145	4
1998	33961	2768	4709	0,139	4
1999	33434	2602	4546	0,136	4
2000	29436	2350	3842	0,131	4
2001	27901	2344	3435	0,123	4
2002	30331	2459	3464	0,114	4
2003	25921	2240	3024	0,117	4
2004	25827	2308	2492	0,096	4
2005	26377	2354	2553	0,097	4
2006	24885	2199	2502	0,101	5
2007	24439	2223	2339	0,096	4

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
2008	24003	2191	2461	0,103	5
2009	25245	2346	2249	0,089	5
2010	25401	2143	2234	0,088	5
2011	25966	2159	2108	0,081	5
2012	25793	2173	2226	0,086	5

The Journal of the American Medical Association

Год	Ссылки	Самоцитирования	Повторные ссылки	Фокус-фактор	Медианный период полураспада цитирования
1992	16383	1672	1889	0,115	4
1993	15229	1679	1785	0,117	4
1994	14988	1746	2011	0,134	4
1995	15464	1777	1943	0,126	4
1996	15290	1557	1766	0,116	4
1997	17411	1748	1772	0,102	4
1998	15720	1725	1863	0,119	4
1999	14951	1553	1817	0,122	4
2000	14770	1485	1855	0,126	4
2001	16079	1388	1691	0,105	4
2002	18069	1669	2108	0,117	5
2003	17338	1698	2314	0,133	4
2004	14948	1346	1871	0,125	5
2005	15857	1255	1660	0,105	5
2006	13982	1241	1648	0,118	4
2007	13190	1035	1440	0,109	5
2008	12960	908	1089	0,084	5
2009	12215	1049	1112	0,091	5
2010	10600	913	1097	0,103	5
2011	10321	791	992	0,096	5
2012	11298	845	991	0,088	5

Научная продукция по открытому доступу: широкомасштабный библиометрический анализ в академическом и научном контексте*

Сандра МИГЕЛЬ
(Sandra MIGUEL)

Институт исследований в области гуманитарных и социальных наук, Национальный университет Ла-Платы, г. Ла-Плата, Аргентина

Эли Франсина Таннури де ОЛИВЕЙРА
(Ely Francina Tannuri de OLIVEIRA)

Мария Клаудия Кабрини ГРАСИО
(Maria Cláudia Cabrini GRÁCIO)

Университет шт. Сан-Паулу, г. Маринья, Бразилия

Это исследование направлено на диахронический анализ широкомасштабной научной продукции по открытому доступу в академическом и научном контексте с целью способствовать знанию и визуализации его основных участников. В качестве метода использовалось библиографическое, дескриптивное и аналитическое изучение вместе с библиометрическими исследованиями, особенно показателей производства, научного сотрудничества и совместной тематической встречаемости. База данных Scopus применялась в качестве источника поиска статей по теме, суммарная совокупность статей составила 1179. Используя программное обеспечение Vindex, для переменных были созданы таблицы частоты использования, программа Rajeek применялась для визуализации сети сотрудничества, а VoSViewer – для создания сети ключевых слов. Что касается результатов, то самыми продуктивными учеными были представители таких стран, как США, Канада, Франция и Испания. Журналы, пользующиеся более высоким влиянием в академическом сообществе, распространяют новое созданное знание. Рассматривалась сеть сотрудничества с несколькими подсетями, в которой соавторами были люди из разных стран. Делается вывод, что это исследование допускает установление тем для дискуссий, которые сопровождают развитие открытого доступа на международном уровне, и можно утверждать, что открытый доступ является одной из новых, появляющихся и пограничных областей библиотековедения и информатики.

ВВЕДЕНИЕ

В течение столетий модель научной коммуникации основывалась на публикации статей сначала в печатных, а затем также и в электронных журналах, где подписная стоимость за доступ к статьям была доминирующей системой распространения и обращения публикаций. Однако за последнее десятилетие эта модель прошла через глубокие изменения, особенно мотивируемые появлением открытого доступа (Open Access - OA) к научному знанию.

Открытый доступ возник раньше этого столетия, частично в результате явления, известного как «кризис в научном издательстве», который стал последствием вы-

соких цен на подписку, сокращения библиотечных бюджетов, последующей отмены подписки и ограничения доступа к статьям для научных сообществ [1,2]. Сформулированный первоначально в Будапештской декларации в декабре 2001 г. [3], а затем несколькими годами позже в Берлине [4] и Бифезде [5], OA предлагает свободный доступ к научным публикациям как альтернативу традиционной модели подписки на журналы и платы за доступ к статьям. Одним из основных принципов OA является разрешение пользователям читать, скачивать, копировать, распространять, печатать, искать или подключаться к полным текстам статей без экономических, юридических или технических барьеров в отличие от самого Интернета. Единственное поставленное условие - позволить авторам контролировать целостность их работы и дать право быть должным образом признанными и цитируемыми.

В недавнем исследовании Бьёрк с соавторами [6] суммировала основные аргументы, поддерживающие

* Перевод Miguel S., Oliveira E.F.T., Grácio M. C. C. Scientific production on open access: A worldwide bibliometric analysis in the academic and scientific context.— <http://www.mdpi.com/2304-6775/4/1/1>

принципы ОА. Первым является увеличение обзримости научных результатов, вносящих вклад в более быстрое продвижение науки. Другим аргументом служит то, что поскольку большая часть науки финансируется общественными фондами, ее результаты должны быть публично доступными. Кроме того, одним из ожидаемых последствий в системе научного издательства является предельное сокращение стоимости публикаций и процессов распространения по сравнению с моделью по подписке.

С самого начала были определены два способа предоставления открытого доступа: золотая дорога, которая предлагает публикацию статей в журналах открытого доступа, и зеленая, предлагающая самоархивирование статей, опубликованных в традиционных подписных журналах, в цифровых хранилищах открытого доступа. Гразиотин, Ванг и Абраамсон [7] детально рассматривают понятия и определения, вовлеченные в ОА.

В 2012 г. Будапештская инициатива открытого доступа [3] еще раз подтвердила принципы ОА, а также два развивающихся способа, указывая, что они не только являются прямыми и эффективными средствами достижения открытого доступа, но также непосредственно доступны академическому сообществу и позволяют аннулировать издательские рынки и избежать законодательных препятствий. Тем не менее, основные издательства повсеместно адаптируют модели открытого доступа, что не гарантирует прямого открытого доступа к статьям во всех случаях, вводя жесткие сроки на доступность в хранилищах, или в других случаях перекладывая стоимость подписки на доступ с помощью авторской платы (гибридные журналы) или платы за подготовку статьи к печати.

Хотя будущее открытого доступа неопределенно, некоторые авторы, такие как Харнад и др. [8], предполагают, что зеленая дорога является единственной альтернативой достижения 100% открытого доступа. Тем не менее, то, «в чем, когда и где» самоархивирование имеет место, является противоречивым [9]. Другие авторы утверждают, что золотой открытый доступ лучше и у него больше потенциала для экономической стабильности, чем у зеленого открытого доступа [10]. Сьюбер [11] предполагает, что оба способа открытого доступа должны дополнять друг друга. Правда заключается в том, что совокупность доступных в открытом доступе статей все еще низкая. В обзоре относительно недавних исследований Бьёрк с соавторами [6] сообщили, что доля статей открытого доступа в журналах, включенных в базы данных Web of Science (WoS) и Scopus, составляет только 23% (с учетом двух способов), хотя уровень их развития отличается в тематических областях и по географическим регионам [12-14].

Важность этого вопроса, получившая широкий размах в разных секторах, отвечающих за развитие науки и научной коммуникации, включая, помимо прочих, авторов, редакторов, библиотекарей, управляющих научной политикой, привела к продолжению работы над исследованиями, составляющими значительный объем научной продукции относительно открытого доступа.

Однако имеется несколько научных статей, проводящих анализ литературы с точки зрения библиометрии, анализирующих международные публикации и носящих междисциплинарный охват. Одна статья содержит исследование структуры и эволюции темы открытого доступа в журналах, включенных в WoS в 2000-2010 гг., на основе анализа коцитирования. Это исследование обнаружило, что открытый доступ являет-

ся зарождающейся областью. Первыми тремя важными статьями являются: статья Лоуренса «Свободная онлайн-доступность потенциально усиливает влияние статьи» (2001 г.), статья Антельмана «Имеют ли статьи открытого доступа большее научное влияние?» (2004 г.) и статья Свон и Брауна «Саморхивирование в открытом доступе: авторское исследование» (2005 г.). Харнад является самым влиятельным автором, а журнал *Journal of the Association for Information Science and Technology* (JASIS/JASIST) – самым часто цитируемым в литературе открытого доступа [15]. Другие исследования проводят обзоры литературы, изучающие взаимосвязь между статусом открытого доступа и подсчетами ссылок научных статей [16].

Учитывая, что тема открытого доступа стала более распространенной с 2002 г., существует ряд исследований, сконцентрированных в мировом масштабе на направлениях и тенденциях развития этой темы, диахронически анализирующих ее присутствие в основном потоке науки, определяющих страны, которые стали ведущими в разработке исследований на эту тему, а также мировую элиту ученых и ядерные научные журналы в распространении этого знания. Современные библиометрические исследования на эту тему посвящены анализу регионов, по областям или дисциплинам, терпя из-за этого неудачу в предоставлении широкомасштабной актуальной визуализации темы.

В этом контексте данное исследование нацелено на диахронический анализ широкомасштабной научной продукции по открытому доступу в академическом и научном контексте, чтобы способствовать знанию и визуализации ее основных участников.

Точнее, это исследование предлагает:

- диахронический анализ присутствия темы открытого доступа в международной научной литературе;
- установление ведущих стран и основных авторов;
- развитие ядерных журналов, распространяющих эту тему;
- обзор сети научного сотрудничества среди наиболее продуктивных авторов;
- определение основных подтем, принадлежащих этой теме, и их совместной встречаемости в анализируемой литературе, через слова, представленные в названиях и рефератах к статьям.

ДАННЫЕ И МЕТОДЫ

Это исследование является библиографическим, дескриптивным и аналитическим с учетом вклада библиометрических исследований, особенно показателей производства, научного сотрудничества и показателей тематической совместной встречаемости.

Традиционно источником информации, используемым для библиометрических исследований на мировом уровне, было совместное множество баз данных, принадлежащих Web of Science (WoS) агентства Thomson Reuters. Одной из их сильных сторон является междисциплинарность и интернациональный характер. Эти базы данных содержат информацию из примерно 12 тыс. научных журналов и стали компетентными мировыми средствами для информационного поиска и исследований, оценивающих науку. Однако появление на рынке базы данных Scopus издательства Elsevier с охватом свыше 20 тыс. журналов помогло дополнить и расширить анализы, взятые из других источников. Уже с первых дней база данных Scopus была предметом анализа [17-19] – сравнений охвата, доступности, практичности и стоимости между ней и WoS [20,21] или сравнения этих двух продуктов с Google Scholar [22,23]. Нам следу-

ет избегать чрезмерного упоминания богатой литературы, касающейся Scopus, и просто заметить, что существуют определенные ограничения [21] наряду с колоссальным усилием, предпринимаемым Elsevier. Все еще существуют пробелы относительно охвата и страны принадлежности в журналах. Несмотря на это, уровни хронологического, географического и тематического охвата, поддерживаемые Scopus, на самом деле достаточно точны для обеспечения релевантных и надежных результатов [13, 24,25].

Используя базу данных Scopus в качестве источника, 2 сентября 2015 г. проводился поиск статей по теме открытого доступа с применением в качестве стратегии следующих терминов в опции Расширенного поиска, названной **Стратегией**:

(TITLE (“open access” OR “open archives” OR “free access”) OR KEY (“open access” OR “open archives” OR “free access”)) AND TITLE-ABS-KEY (scholarly OR publishing OR scientific OR libraries OR repositories OR academic) AND DOCTYPE(ar) AND PUBYEAR < 2015 AND PUBYEAR >1900,

где:

TITLE: поиск в названии области

KEY: поиск в области ключевого слова

TITLE-AS-KEY: поиск в областях названия, реферата и ключевых слов одновременно

DOCTYPE(ar): ограничения поиска по типу документа «Статьи»

PUBYEAR< AND PUBYEAR>: ограничения поиска периодом времени в соответствии с годом издания.

Таким образом, эти термины позволяли поиск статей, в котором один из терминов «open access» или «open archives» или «free access» появляется в названии или среди ключевых слов. Для придания смысла этим терминам в контексте научной публикации данный поиск был ограничен наличием терминов *образовательный* или *издание* или *научный* или *библиотеки* или *хранилища*

или *академический* в названии, реферате или ключевых словах. На основе сравнения мы выявили сходство этой поисковой стратегии со стратегией, предложенной методом PICO, методологией, используемой в неклиническом исследовании, в котором P обозначает популяцию, I – интерес, а Co – контекст [26]. В этом исследовании P относится к статьям по теме открытого доступа, представленной в поисковой стратегии терминами TITLE и KEY. Таким образом, I представляет интерес в знании диахронической эволюции открытого доступа в основном потоке науки. Наконец, Co представлено терминами TITLE-ABS-KEY, DOCTYPE(ar) и PUBYEAR. Поиск был отфильтрован, исключая статьи, опубликованные в 2015 г., и те статьи, которые содержали термин «free access», но были не по данной теме. В итоге массив составил 1179 статей. Во всех других найденных статьях отмечалась исследуемая тема. Это множество было экспортировано в формат научно-информационной системы с отбором следующих областей: автор, название, год, источник названия, принадлежность, реферат и ключевые слова.

Далее, используя программное обеспечение Vībexcel, были построены таблицы частоты для переменных автор, год и журнал. Для диахронического анализа распространения темы по всему миру таблица частоты для года издания была конвертирована в линейную диаграмму с помощью программного обеспечения Excel.

Для создания табл. 1, относящейся к странам происхождения авторов, а также к анализу наиболее повторяющихся ключевых слов, использовался Отчет Scopus.

Для визуализации признания самых продуктивных авторов (табл. 2) из исследуемого сообщества было проанализировано общее число статей, в котором цитировался каждый автор, среди найденных статей. По отношению к каждому автору этот поиск ссылок проводился с добавлением поискового термина REFAUTH (имя автора) в опцию Расширенного поиска **Стратегии**.

Таблица 1

Страна	Число статей	%
США	358	30,4
Великобритания	158	13,4
Испания	70	5,9
Германия	67	5,7
Канада	58	4,9
Индия	54	4,6
Австралия	38	3,2
Франция	33	2,8
Нидерланды	29	2,5
Италия	26	2,2
Бразилия	26	2,2
Китай	21	1,8
Финляндия	21	1,8
Австрия	20	1,7
Швейцария	20	1,7
Бельгия	17	1,4
Греция	17	1,4
Нигерия	15	1,3
Иран	14	1,2
Дания	14	1,2
Южная Африка	14	1,2
Швеция	13	1,1

Таблица 2

Ученый (Страна)	Число статей
Björk, Bo-Christer (Финляндия)	18
Xia, Jingfeng (США)	10
Davis, Philip M. (США)	8
Harnad, Stevan (Канада)	8
Schopf, Joachim (Франция)	7
Solomon, David J. (США)	7
Abadal, Ernest (Испания)	6
Bauer, Bruno (Австрия)	6
Laakso, Mikael (Финляндия)	6
Melero, Remedios (Испания)	6
Moskovkin, Vladimir M. (Россия)	6
Mukherjee, Bhaskar M. (Индия)	6
Abrizah, Abdullah (Малайзия)	5
Beall, Jeffrey (США)	5
Jacsó, Péter (США)	5
Kousha, Kayvan (Иран)	5
McGrath, Mike (Великобритания)	5
Nicholas, David (Великобритания)	5
Oppenheim, Charles B. (Великобритания)	5
Rodrigues, Rosângela S. (Бразилия)	5
Sánchez Tarragó, Nancy (Куба)	5
Willinsky, John (США)	5

Таблица 3

Журнал (Страна издания)	Число статей	%	Квартиль	Дорога
Serials Librarian (Великобритания)	34	2,9	Q2 Q1	Зеленая
Learned Publishing (Великобритания)	23	2,0	Q1	Голубая
J. Am. Soc. Inf. Sci. Technol (США)	23	2,0	Q1	Зеленая
Information Services and Use (Нидерланды)	22	1,9	Q3	Зеленая
Serials Review (Великобритания)	21	1,8	Q2	Зеленая
El Profesional de la Información (Испания)	20	1,7	Q2	Голубая
Insights (Великобритания)	20	1,7	Q3	—*
OCLC Systems & Services: International digital library perspectives (Великобритания)	19	1,6	Q3	Зеленая
Journal of Academic Librarianship (Нидерланды)	18	1,5	Q1	Зеленая
First Monday (США)	18	1,5	Q2	Голубая
VOEB-Mitteilungen (Австрия)	16	1,4	Q1	-
Online Information Review (Великобритания)	16	1,4	Q2	Зеленая
D-Lib Magazine (США)	16	1,4	Q1	Голубая
TripleC (Австрия)	14	1,2	Q2	Зеленая
Scientometrics (Нидерланды)	14	1,2	Q1	Зеленая
Interlending and Document Supply (Великобритания)	14	1,2	Q2	Зеленая
LIBER Quarterly (Нидерланды)	13	1,1	Q3	Зеленая
Information Research (Великобритания)	12	1,0	Q3	Голубая

* Журнал не найден в портале SHERPA/ROMEO

Что касается журналов в табл. 3, то для записи страны издания журнала, а также принадлежащего каждому журналу квартиль использовался портал SCImagoJR. Для установления формы открытого доступа этих журналов применялись метаданные открытого архивирования портала SHERPA/RoMEO-Rights, которые позволяют визуализацию типа дороги открытого доступа по журналу или издательству, категоризируя политику архивирования: зеленая (можно архивировать препринты и постпринты или издательскую версию/ PDF), голубая (можно архивировать постпринты, т.е. окончательный

вариант после рецензирования, или издательскую версию /PDF), желтая (можно архивировать препринты, т.е. вариант до рецензирования) и белая (архивирование, формально не поддерживаемое).

Для создания сети соавторства самых продуктивных ученых применялась БД Scopus в целях поиска их соавторов. Для этого поиска термин AUTH (имя автора) был добавлен в опцию Расширенного поиска **Стратегии**.

Для множества соавторов, найденных для каждого самого продуктивного ученого из табл. 2, была создана сеть соавторства соавторов с по крайней мере двумя

статьями, опубликованными с данным ученым. Этот критерий выбран благодаря тому факту, что имеется 101 соавтор, принадлежащий только одной статье среди всего множества, что затруднит визуализацию консолидированного сотрудничества по этой теме. Для создания этой сети использовалось программное обеспечение Pajek.

Программное обеспечение VoSViewer применялось для построения и визуализации сети совместной встречаемости ключевых слов с целью оценить их совместную встречаемость в названиях и рефератах. Всего было извлечено 284 разных термина, с которыми построена сеть совместной встречаемости. Было решено оставить ключевые слова так, как они даны у авторов, чтобы определить различия в словаре, выявить изменения и преобладание. Карты были созданы с использованием методов отображения и кластеризации VoS с цветовым различением кластеров.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Эволюция темы открытого доступа в международной научной литературе

На рис. 1 показано направление развития этой темы в мире от начала дискуссий вплоть до 2014 г., когда наблюдается восходящая кривая, с несколькими годами, отвечающими за отклонения. К 2012 г. научная литература по данному предмету демонстрирует небольшое сокращение.

Как упоминалось во Введении, в декабре 2001 г. конференция Института открытого общества состоялась в Будапеште, когда термин открытый доступ был оформлен и формализован. Затем, в январе 2002 г., Будапештская декларация инициативы открытого доступа (Budapest Open Access Initiative - BOAI) установила принципы, стратегии, правила и требования, связанные с открытым доступом в научной литературе.

Эти факты объясняют первую точку отклонения кривой в 2002 г., получившего развитие в 2006 г. с появлением PLoS One. В 2008 г. покупка издательства открытого доступа BioMed Central издательством Springer,

одним из лидеров рынка в академическом журнальном издательстве, укрепила модель открытого доступа [27].

С тех пор публикации по этой теме показывают резкий скачок кривой за исключением 2001 г., когда отмечается значительный спад, что заслуживает более подробного изучения с целью определить его причины.

Страны и основные авторы темы открытого доступа в международной научной литературе

Табл. 1 показывает 22 основные производящие страны из всех 174 стран, публикующих работы на тему открытого доступа. Каждая страна дала не менее 13 статей за период, соответствующий примерно 1% всего производства статей по теме и их долей.

Отмечается, что представленные в табл.1 страны в общем являются самыми важными в научном мире с некоторыми исключениями, такими как Япония, которая согласно portalу SCImago занимает пятое место в мировом рейтинге и не находится среди основных стран, производящих статьи по открытому доступу, а также Южная Корея, Российская Федерация и Тайвань. С другой стороны, Испания, Нидерланды и Бразилия занимают верхние ранги в научной литературе по открытому доступу относительно тех стран, которые заняты в мировой науке; предполагается, что эти страны, особенно Испания, предпринимают значительные усилия в развитии анализируемой темы.

Что касается направлений стратегий открытого доступа, то акцентируется, что США отдают предпочтение зеленой дороге. В 2013 г. Закон о справедливом доступе к исследованиям в науке и технике (FASTR) требовал открытого доступа к статьям из основных федеральных агентств, финансируемых за счет хранилищ. В их научном производстве самыми повторяющимися ключевыми словами, индексирующими статьи, помимо прочих являются: статьи (62), доступ к информации (48), библиометрия и ключевые слова, связанные с ее процедурами (31), журналы (30), база данных (30) и хранилища (22). Акцентируя внимание на ключевом слове «статьи», отмечаем выравнивание между выпущенной научной литературой и привилегированной стратегией открытого доступа, принятой страной (зеленая дорога).

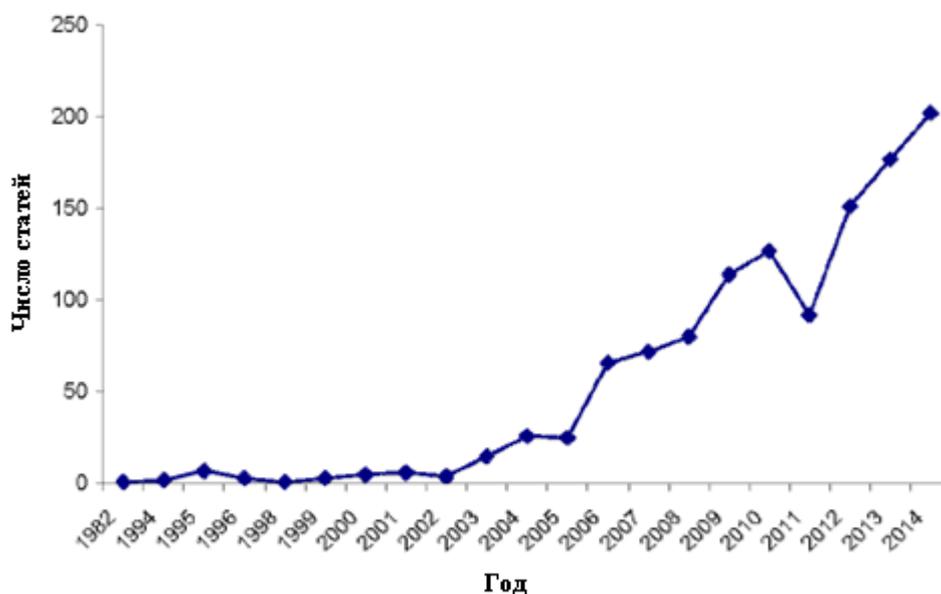


Рис. 1. Эволюция темы «открытого доступа» в Scopus

Великобритания через Научный совет поощряет открытый доступ с помощью золотой дороги путем широкого стимулирования издателей. В ее научном производстве наблюдается широкое рассеяние ключевых слов для индексирования статей, проявляющееся следующим образом: статьи (31), доступ к информации (23) и журнал (15).

В Испании Закон о науке, технологии и инновациях от 14/2011 г. в статье 37, рассматривающей распространение в открытом доступе, объясняет, что государственные должностные лица испанской системы науки, технологии и инновации должны поощрять развитие хранилищ открытого доступа для публикаций, принадлежащих их ученым, и установить системы, позволяющие связать их с подобными инициативами на национальном и международном уровнях. Ее научная продукция выделяет следующие ключевые слова в статьях: хранилище (21), доступ к информации (15), журналы (22), статьи (14) и библиометрия и ключевые слова, связанные с ее процедурами (9).

В Нидерландах двадцать институтов, включая Голландскую королевскую академию искусств и наук (KNAW), Ассоциацию университетов Голландии (VNSU) и Голландскую ассоциацию университетов прикладных наук (Vereniging Hogescholen), подписали Берлинскую декларацию [28]. Голландское правительство намерено полностью перейти на открытый доступ с помощью золотой дороги в течение примерно 10 лет, т.е. к 2024 г. Наряду с этим предложением Голландская организация по научному исследованию (NOW), основной финансирующей научного исследования в голландских университетах и институтах, первоначально адаптирует программу золотого открытого доступа для научных фондов, а где это невозможно, поддерживает программу зеленого открытого доступа [29]. Следующие ключевые слова стоят особняком в научном производстве этой страны: журналы (5), статья (4) и хранилище (2) с широким рассеянием терминов индексированных статей.

В Бразилии Законопроект 387/2011 основан на зеленой модели. По этому проекту учреждения высшего образования и другие научные единицы должны построить институциональные хранилища открытого доступа с использованием стандартизированных международных метаданных. Следующими ключевыми словами, индексированными статьи, являются: журналы (6), доступ к информации (4), хранилище (3) и статьи (2).

В отношении авторов, отвечающих за 1179 статей, мы нашли в общем 2407 авторов, что в среднем соответствует 2,04 автора на статью. Стоит отметить, что 2108 авторов создали одну единственную статью (87,6%), это наводит на мысль о широком рассеянии ученых, производящих научную литературу по данной теме.

В табл. 2 представлено 22 наиболее продуктивных ученых, расположенных в нисходящем порядке качества публикуемых статей с публикацией по крайней мере пяти статей по теме. Этот критерий был принят, так как 50 авторов опубликовали четыре статьи, образуя широкую дисперсию.

Первоначально выявлено, что 22 ученых, представленных в табл.2, отвечали вместе за 132 (11,2 %) статьи из общего числа проанализированных статей, все они были опубликованы, начиная с 2003 г. Таким образом, 34 статьи, включающие период 1982-2002 гг., были от менее количественно выраженных авторов, учитывая, что индивидуально они опубликовали менее четырех статей.

Исходя из табл. 2, найдены ученые из разных стран, но особенно из США, Великобритании, Испании и

Финляндии. Среди ученых с самым высоким числом статей обнаружены первооткрыватели и последователи темы открытого доступа, такие как Стивен Харнад, фигурирующий в исследовании авторов [15].

Среди 22 ученых, представленных в табл. 2, 13 имеют значительное влияние на множество из 1179 анализируемых статей, будучи цитируемыми по крайней мере в 1% этих статей, а именно: Харнад (20,4%), Ошпенхайм (12,4%), Бьёрк (10,6%), Дэвис (8,3%), Виллински (7,8%), Лааксо (6,6%), Николас (5,9%), Ся (3,6%), Соломон (3,6%), Коупа (3,1%), Билл (2%), Джакó (1,6%) и Мелеро (1,4%).

С. Харнад, защитник и сторонник зеленой дороги, из Университета Квебека в Монреале, выступал за открытый доступ в 1990-х гг., предлагая, чтобы вся научная литература была доступна в Интернете бесплатно и неограниченно, а также указывая названия стратегий – зеленая дорога, золотая дорога и самоархивирование [30]. Его статьи, найденные в ходе этого исследования, опубликованы в период 2005-2013 гг. индивидуально или в соавторстве с учеными из Канады, Великобритании и Франции. Он, в частности, изучает вопросы, связанные с когнитивной наукой и открытым доступом, которые дают основу для исследования многим ученым из разных стран и со всех континентов. В рамках анализируемой литературы он был самым цитируемым автором (в 240 из 1179 анализируемых статей) в научной продукции ученых из 41 разной страны, особенно США, Великобритании, Испании, Индии, Канады, Финляндии, Германии, Ирана и Бразилии, писавших на семи разных языках, в частности английском, испанском и португальском. Этот результат согласуется с тем, что нашли авторы работы [15], определяя Харнада как самого влиятельного автора по данной теме.

Чарльз Б. Ошпенхайм, из Лондонской бизнес-школы (Cass Business School, London), эксперт в области авторского права, защиты данных, свободы информации, открытого доступа и других аспектов распространения исследования, библиометрии и оценки исследований, с пятью статьями, опубликованными в период 2003 – 2010 гг., является вторым самым цитируемым автором в анализируемой литературе (цитируется в 146 статьях), на которого ссылаются ученые из 37 разных стран, особенно из США, Великобритании, Испании, Индии, Канады, Аргентины, Японии и Финляндии, в статьях, опубликованных на шести разных языках, главным образом английском и испанском.

Бо-Кристер Бьёрк, ученый отделения менеджмента и организации Школы экономики в Хельсинки (Hansen School of Economics) и сторонник модели научной коммуникации, названной глобальной распределительной информационной системой (2007 г.), был автором с наиболее высоким числом опубликованных статей, с почти в два раза большим числом по сравнению со вторым самым продуктивным автором. Его публикации, включенные в это исследование, регулярно распространяются в 2004-2014 гг. и публикуются отдельно или входят в элиту переднего края научного исследования в изучаемой теме.

Филип М. Дэвис, ученый фирмы Phil Davis Consulting, Итака, совсем недавно (2008-2013 гг.) опубликовал восемь представленных в табл. 2 статей индивидуально или в соавторстве с учеными из его собственной страны, изучая в частности тему открытого доступа с библиометрическими анализами, особенно анализом цитирования. Теоретически на него ссылались 98 статей, опубликованных авторами из разных стран (28), особенно США, Великобритании, Испании и Канады.

Джон Виллински, ученый из Стэнфордского университета (Высшая школа образования, Пало-Альто), издал пять статей индивидуально или в соавторстве с учеными из его собственной страны и Канады в период 2003-2013 гг., особенно рассматривая открытый доступ в области медицины. Его статьи содержат теоретические ссылки на 92 из 1179 проанализированных статей, опубликованных учеными из 27 разных стран, главным образом США, Великобритании, Бразилии и Испании, и на пяти разных языках, в основном английском, португальском и испанском.

Микаэль Лааксо из отделения менеджмента и организации, Шведская школа экономики (Хельсинки), опубликовал свои статьи недавно (2010-2014 гг.), шесть из них в соавторстве с учеными из пяти различных стран, а также с учеными из своей страны. Статьи особенно обсуждают как мета-теоретические исследования по открытому доступу, так и вопросы, связанные со стратегией зеленой дороги. Его статьи формируют теоретические ссылки на 78 из 1179 проанализированных статей, изданных учеными из 27 различных стран, главным образом США, Великобритании, Испании, Канады, Финляндии и Индии, и опубликованных на семи разных языках, в частности английском и испанском.

Дэвид Николас из Центра информационного поведения и оценки исследования (Школа библиотечных, архивных и информационных исследований Лондонского университетского колледжа) опубликовал все свои статьи в соавторстве с учеными из США и Ирана, а также с учеными его собственной страны в области социальных и информационных наук в период 2005-2009 гг. Эти статьи формируют теоретические ссылки на 70 из 1179 проанализированных статей, изданных исследователями из 24 различных стран, в частности из США, Великобритании и Австралии, и опубликованных на четырех разных языках, особенно английском и испанском.

Цзинфен Ся из отделения по библиотечному делу и информатике (Университет шт. Индиана, США) показал значительное научное производство статей по данной теме, распространенной в 2009-2012 гг., в частности связанной с хранилищами и анализом влияния научной продукции по открытому доступу через ее цитирование. Он разработал исследование индивидуально и в соавторстве с учеными из своей страны. Среди 1179 проанализированных статей он был процитирован в 43 статьях, особенно в области социальных наук и вычислительной техники, изданных на английском языке авторами из США, Испании, Канады и Германии.

Дэвид Дж. Соломон из Медицинского колледжа (Мичиганский государственный университет) опубликовал свои статьи под индивидуальным авторством или в сотрудничестве с финскими учеными Бьерком и Лааксо, особенно изучая вопросы, связанные с журналами открытого доступа в медицине и мета-теоретическими исследованиями, основанными на библиометрических исследованиях, в 2006-2014 гг. Он был процитирован в 42 статьях, как правило учеными из США, Канады, Финляндии, Испании и Великобритании, а также из 26 других стран, изданных главным образом на английском и испанском языках.

Кейвин Коупа из отделения по библиотековедению и информатике (Тегеранский университет) опубликовал свои статьи как в индивидуальном авторстве, так и в соавторстве с учеными из Великобритании (Майк Тэлволл) и Испании в 2007-2014 гг., в области социальных и вычислительных наук, особенно обращаясь к проблеме влияния и цитирования публикаций открытого доступа. Эта научная продукция заложила основу для 37

статей из множества проанализированных статей, созданных учеными из семи стран, главным образом США, Канады, Индии и Ирана.

Джеффри Билл из Университета шт. Колорадо в Денвере индивидуально опубликовал найденные в ходе этого исследования статьи в 2009-2013 гг., в которых он представляет размышления, дискуссии и протирочения, возникающие вокруг открытого доступа, среди них и вопрос о завышенных грабительских ценах издателей и о финансировании открытого доступа через золотую дорогу. На эту научную продукцию ссылались в 23 статьях, изданных 4 учеными из 10 стран, включая США, Великобританию и Финляндию.

Петер Джакó из отделения вычислительной техники и информационных систем (Гавайский университет, Гавайи) опубликовал статьи в период 2003-2010 гг. в индивидуальном авторстве, обсуждая технические и методологические аспекты индексирования и научно-информационного поиска и библиометрического анализ по теме открытого доступа. На эту научную продукцию ссылались в 19 статьях, изданных авторами из 13 стран, в частности Индии, Великобритании и США.

Ремедиос Мелеро, испанский ученый Института агрохимии и технологии пищевых продуктов, Валенсия, опубликовала свои статьи в период 2009-2014 г. в индивидуальном авторстве, а также в соавторстве с другими испанскими учеными в области социальных наук и вычислительной техники, главным образом на испанском языке. Ее научная продукция дала материал для 17 статей проанализированного массива ученых из Испании, Аргентины и Бразилии, Мексики и Малайзии, изданных на английском и испанском языках.

Стоит отметить, что девять представленных в табл. 2 ученых с меньшим влиянием (цитируемых менее 1% статей) в проанализированной литературе вообще опубликовали свои статьи совсем недавно, начиная с 2009 г., а большинство из них изданы в 2014 г. – последнем году проанализированного периода, это Дж. Шопфель (2005-2014 гг.), Е. Абадал (2009-2014 гг.), Б. Бауэр (2009-2014 гг.), В. Московкин (2008-2014 гг.), Б. Макхерджи (2007-2014 гг.), А. Абризах (2009-2013 гг.), М. МакГрат (2012-2013 гг.), Р. Родригес (2010-2014 гг.), Н. Санчес-Тарраго (2007-2012 гг.). Таким образом, одна из возможных гипотез низкого влияния – это факт того, что научное сообщество менее вероятно обратится к этим публикациям, так как среди авторов с самым большим влиянием их статьи вообще были распространены в предыдущие годы.

Другая возможная гипотеза низкого влияния касается того, что эти ученые происходят не из англо-говорящих стран и издаются индивидуально или в соавторстве только с коллегами из их собственных или соседних стран, а также не англоязычных стран, таких как Германия (Бауэр), Сербия и Россия (Московкин), Бразилия (Абадал), Малайзия и Иран (Абризах) и Испания (Родригес и Санчес-Тарраго). Эта характеристика совпадает с наблюдением нескольких ученых по библиометрии, которые утверждают, что существует тенденция: чем выше международное научное сотрудничество, тем выше влияние изданной литературы [31,32].

Ядерные журналы по теме открытого доступа в международной научной литературе

Табл. 3 показывает ядерные журналы по изучаемой теме, основываясь на выборке 18 названий из 488 журналов, которые распространяли найденные статьи. Каждый из журналов издал по крайней мере 1% этих 1179 статей, насчитав в сумме 333 статьи, что соответствует 28,2 % публикаций.

Стоит отметить, что это ядро из 18 журналов издается группой, состоящей из пяти стран; восемь имеют британское происхождение (Великобритания). Нидерланды, страна крупных коммерческих издателей, распространяющих научные публикации, является страной происхождения четырех журналов, все заиндексированы в области библиотечного дела и информатики. США, самый большой производитель статей по рассматриваемой теме (табл. 1), являются страной происхождения трех журналов.

Что касается квартилей, то 67 % журналов принадлежат Q1 и Q2 в социальных науках, предполагая, что журналы с более высоким влиянием в научном сообществе распространяют новое созданное по теме знание.

Большинство журналов (11) категоризировано на портале SHEPRA/Romeo под зеленой дорожкой политики архивирования, то есть разрешая архивирование препринта и постпринта или версии издателя/PDF, и значительную часть (пять) из них под голубой дорожкой политики архивирования, которая позволяет архивирование постпринта (т.е. окончательный вариант после рецензирования) или версии издателя/PDF. Два не размещенных на этом портале журнала являются журналами золотой дорожки. Интересно отметить, что издание журналов в открытом доступе в основном не является золотой дорожкой. Недавнее исследование, проводимое Эннасом и Ди Гуардо [33] обнаружило, что журналы золотой дорожки БД Scopus, входящие в верхний ранг SCImagoJournal & Country Rank (SJR), имеют тенденцию требовать плату за публикацию; только издатели, расположенные в США и Великобритании, положительно относятся к этому ранжированию и только издание журналов на английском языке положительно относится к высоким ранжированиям.

Таким образом, наблюдается беспокойство авторов относительно согласования исследуемой темы и типа журнала, выбранного для социализации нового, созданного по этой теме знания.

Сеть научного сотрудничества среди самых продуктивных авторов

Сеть научного сотрудничества самых продуктивных исследователей по теме вместе с соавторами, с которыми ученые опубликовали в соавторстве более одной статьи, представлена на рис. 2. 22 автора, около имен

которых не стоит буква «а», соответствуют самым продуктивным по теме, а толщина сегментов представляет частоту соавторства между двумя авторами.

Сеть фрагментирована на небольшие подсети и не составляет органичное целое. Автор Б.-К. Бьёрк (Финляндия) выявлен в самой большой подсети, как самый продуктивный, имеющий соавторство с пятью учеными, двое из которых принадлежат группе самых продуктивных авторов, а именно: П. Лааксо (Финляндия), с кем он имеет пять соавторств – самое большое число соавторств, и Д. Дж. Соломон (США), с кем он имеет четыре соавторства.

Также отмечается, что автор Б.-К. Бьёрк представляет наибольшую степень центральности, равную пяти, так как он имеет наибольшее число связей, т. е. с П. Веллингом (Финляндия), Т. Хелландом (Финляндия) и З. Тюрком (Словения), вместе с двумя самыми продуктивными авторами, которые упоминались раньше. Рассматривая географические расстояния между происхождением этих ученых, стоит подчеркнуть, что автор Б.-К. Бьёрк имеет широкие связи, и эта подсеть может быть индикатором более плотного создания сети.

Вторая отмеченная подсеть – триада, состоящая из двух самых продуктивных ученых, Р. М. Мелеро (Испания) и Э. Абадал (Испания), и коллеги Ф. Абад-Гарсиа (Испания). Испания с самого начала принимает золотой открытый доступ.

Хотя сотрудничество между испанскими и бразильскими учеными не отражено в сети, интересно рассмотреть недавнюю апрельскую публикацию 2014 г., Росанжелы Родригес и Эрнеста Абадала, которые приводят краткий обзор модели издательства открытого доступа в Бразилии и Испании. Авторы проводят анализ журналов этих стран, которые отражены в Web of Science (WoS) и Scopus, в соответствии с некоторыми аспектами формата публикации – онлайн или печатный, типа доступа – либо открытый, либо по подписке, и используемой технологической платформы.

Пять других небольших подсетей стоят особняком: с П. М. Дэвисом (США) в центре подсети и соавторами Д. Конолли (США), Дж. Г. Бутом (США), Д. Х. Саймоном (США) и Б. В. Левиштейном (США), это типично американская подсеть, которая выделяет зеленую дорожку.

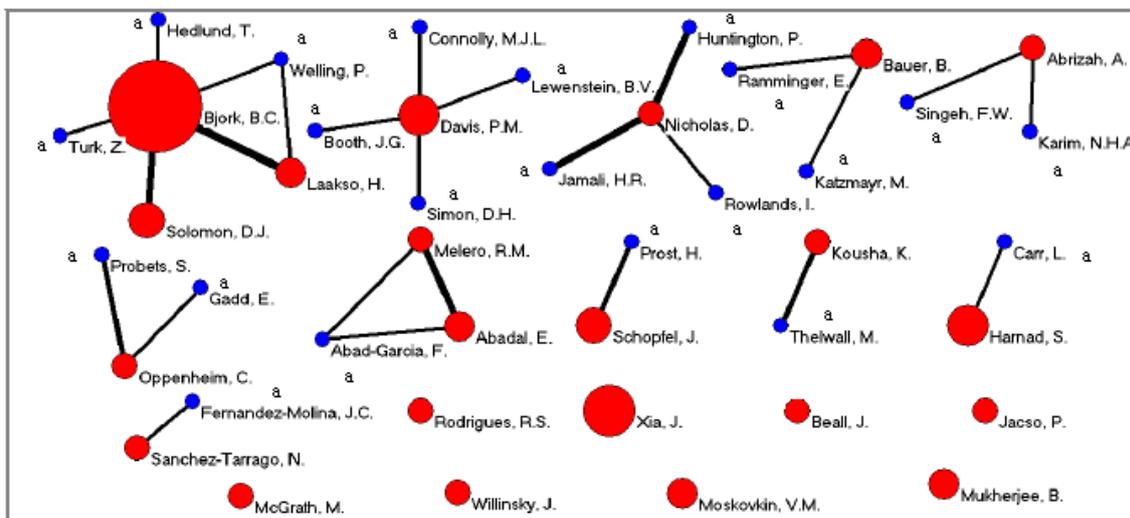


Рис. 2. Сеть сотрудничества среди 22 самых продуктивных авторов, которые имели по крайней мере два соавторства

Д. Николас (Великобритания) имеет четыре соавторства с Х. Р. Джамали (Иран) и четыре с П. Хантингтоном (Великобритания), определенные самыми сильными долями, и с Х. Роулэндсом (Великобритания). Стоит отметить однородность этой группы, правда один ученый происходит из азиатской страны.

Представлены небольшие подсети, сосредоточенные на Б. Бауэре (Австрия) и его двух коллегах, Е. Камингер (Австрия) и М. Катцмаур (Австрия), другая подсеть сосредоточенная на А. Абризахе (Малайзия) и его двух коллегах, Ф. У. Синджей (Малайзия) и Н. Х. А. Карим (Малайзия), и, наконец, Ч. Оппенхайм (Великобритания) и двое его коллег, Е. Гаддс (Великобритания) и С. Пробетс (Великобритания).

Пока еще есть четыре диады, которые в каждом случае включали самых продуктивных авторов в соавторство с другим ученым; восемь из наиболее продуктивных авторов идут отдельно. Замечено, что среди них двое, Джакó и Билл, работают индивидуально, а МакГрат, Московкин, Родригес, Макхерджи, Виллински и Ся имели соавторства, но с авторами одной единственной работы. Поэтому эти последние авторы не рассмотрены в подсети. В качестве выводов сетевого анализа отмечено, что есть несколько подсетей, где соавторы происходят из разных стран, такие как первая и наибольшая подсеть сконцентрированная на Б.-К. Бьёрке, и небольшая подсеть, сосредоточенная на Д. Николасе. Другие являются небольшими подсетями, состоящими из соавторов, происходящих из одной страны, и показывающими зарождающуюся сеть сотрудничества, с подсетями, которые едва экстраполируют к пределам их собственных стран, оправданными недавним расширением вопроса и ограниченностью политики, поощряющей открытый доступ.

Основные подтемы, имеющие отношение к открытому доступу

В течение анализируемого периода выявлены следующие подтемы из зарегистрированных ключевых слов: распространение и информационный доступ (221), статьи (192), журналы (155), хранилища (151), библиотеки (122), библиометрия (74) – в основном анализ цитирования, анализ импакт фактора и научного сотрудничества - и база данных (66). Публикации, рассматривающие стратегию, названную «зеленой дорогой», где открытый доступ предоставляется через самоархивирование автором научных исследований в хранилищах, могут быть представлены условиями статей, хранилищ и библиотеки. С другой стороны, исследования, касающиеся стратегии, названной «золотой дорогой», относятся к научным журналам открытого доступа и представлены ключевыми словами журналов и импакт фактором среди других аспектов, ассоциирующихся с проанализированными статьями.

Карта совместной встречаемости слов в названиях и рефератах статей, приведенная на рис. 3, показывает основные ядерные темы, представляющие международную литературу по открытому доступу в анализируемом периоде. Можно визуализировать пять кластеров. В каждом кластере каждый узел отмечается словом или предложением, извлеченным из названий и рефератов. Его размер – показатель его относительной важности в множестве сетей, а близость с другими узлами служит показателем сходства, основанного на частоте совместного появления слов или предложений в одном и том же названии или реферате. Линии, соединяющие узлы, представляют взаимосвязь между ними.

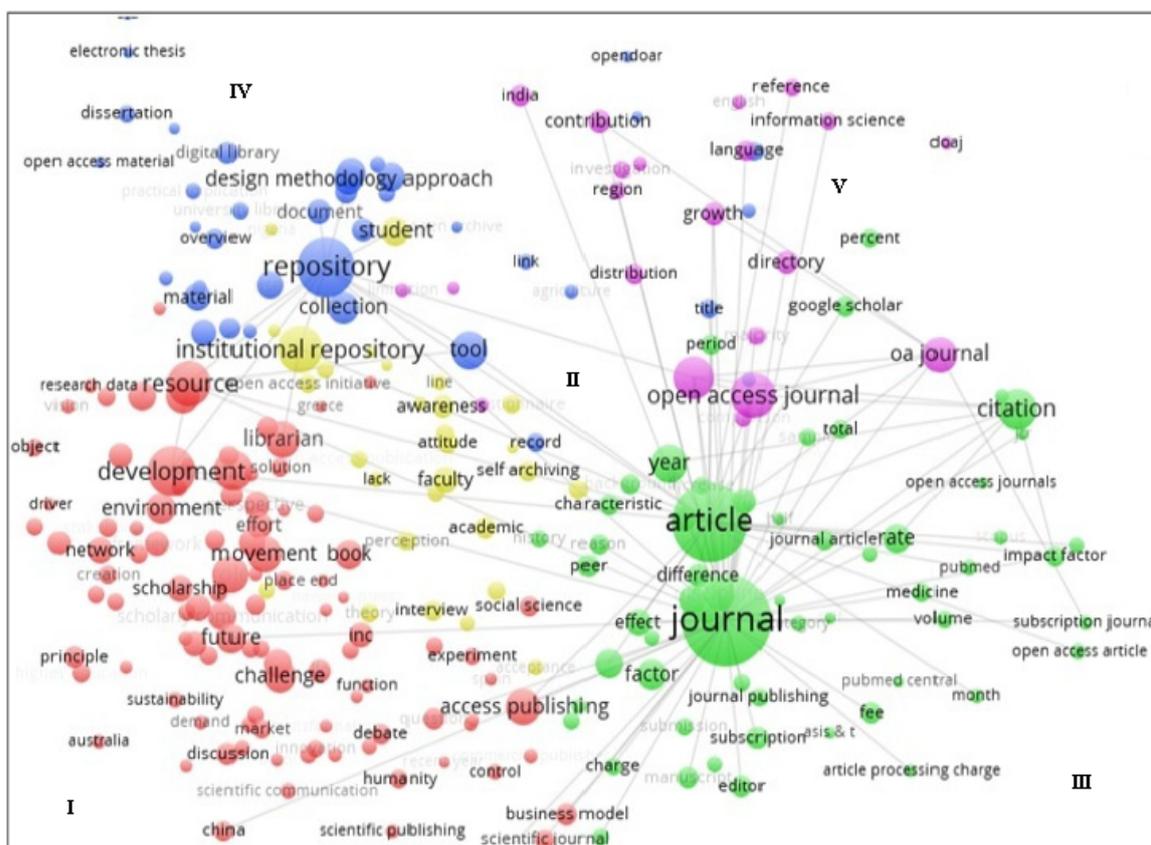


Рис. 3. Сеть совместной встречаемости слов в названиях и рефератах

Кластер 1 идентифицирован на рис. 3 как область I, группа изучает подход к теме открытого доступа с точки зрения изменений в модели научной коммуникации. Кластер 2, представленный как область II, включает исследования об отношениях и восприятии открытого доступа научным сообществом. Кластер 3 – область III представляет ядро статей о золотой дороге открытого доступа, а также ее воздействия на цитирование и влияние журналов открытого доступа. Кластер 4 (область IV) относится к исследованиям зеленой дороги открытого доступа, в которых центральную роль занимают хранилища и библиотеки. Кластер 5 (область V) включает исследования, касающиеся роста открытого доступа в директориях журналов (DOAJ) и в хранилищах (например, OpenDOAR), которые являются выражением уровня развития.

ЗАКЛЮЧЕНИЕ

Это исследование стремилось охарактеризовать самые важные моменты открытого доступа на международном уровне относительно его присутствия в мировой научной литературе, с помощью временного анализа, чтобы установить его научную элиту и ядерные журналы по распространению нового, созданного по этой теме знания. Оно также стремилось представить сеть соавторства, выявленного по данной теме, и идентифицировать основные подтемы и их совместную встречаемость в анализируемой литературе с помощью индексирования статей ключевыми словами.

Важно обратить внимание, что движение открытого доступа считает знание товаром, публично доступным всем гражданам, чтобы предоставить обществу новое знание, происходящее из нового исследования.

Следовательно, предполагается, что для развивающихся стран эта модель важна для развития равенства, а также увеличения способности ученых продвигаться в исследовании через бесплатный доступ к недавно созданной науке.

Отмечалась важная роль конференции, созванной Институтом открытого общества, в декабре 2011 г., в Будапеште, с последующим принятием декларации ВОАИ, в которой были выделены принципы, стратегии, правила и требования, связанные с открытым доступом к научной литературе для немедленного роста научной продукции по теме.

Если с одной стороны открытый доступ подходит для распределения с помощью демократического способа нового созданного знания, то было обнаружено, что основными странами, работающими над темой, являются США, Великобритания, Испания и Германия, т.е. страны, принадлежащие к господствующей тенденции науки, но также и такие страны, как Индия, Бразилия, Нигерия и Иран, которые считаются развивающимися, занимают видное место в ранжировании наиболее продуктивных стран, занятых указанной темой.

Такое видное место указанных выше развивающихся стран связано с самыми продуктивными учеными, которые происходили из таких стран, как США, Канада, Франция и Испания, таких как Российская Федерация, Индия, Бразилия и Малайзия, более отдаленные страны. Таким образом, согласно этим данным, считается, что открытый доступ важен для стран в любой стадии развития, поскольку это предполагает рост и демократизацию знания.

Что касается журналов, то было замечено беспокойство авторов относительно социализации нового знания, полученного по рассматриваемой теме в журналах,

издательская политика которых позволяет архивировать статьи и выбирать или зеленую, или золотую дорогу. Выявлено, что большинство журналов, относящихся к ядерным журналам, имеет британское происхождение, принадлежат к Q1 и Q2. Кроме того, журналы со значительным влиянием на научное сообщество распространяют новое, сгенерированное по теме знание.

Наблюдалась сеть сотрудничества с несколькими подсетями, где соавторы из разных стран и несколько небольших подсетей, состоящих из соавторов одной и той же страны; это показывает, что блокировка несколько превосходит границы их собственных стран. Возможно недавний рост темы и не полностью раскрываемая политика объясняют ограниченное международное соавторство.

Кластеры, связанные с совместной встречаемостью ключевых слов, демонстрируют пять тематических групп исследования: открытый доступ с точки зрения изменения в модели научной коммуникации; отношение и восприятие открытого доступа научным сообществом; золотая дорога открытого доступа, а также ее воздействия на цитирование и влияние журналов с открытым доступом; зеленая дорога открытого доступа, чьи хранилища и библиотеки занимают центральную роль; рост открытого доступа в директориях журналов (DOAJ) и рост хранилищ (например, OpenDoar), которые являются выражением признания развития.

Это исследование, основанное на соответствующей научной литературе, не только позволяет идентифицировать темы дискуссий, которые отмечают развитие открытого доступа на международном уровне, но также и, с точки зрения анализа профилей авторов и самых распространенных журналов, дают возможность заявить, что открытый доступ – одна из новых появляющихся и пограничных областей библиотекостроения и информатики.

Однако мы признаем, что несколько журналов, рассматривающих открытый доступ в качестве темы, еще не заиндексированы в Scopus. Таким образом, мы предлагаем дальнейшие исследования, которые будут проведены с использованием Google Scholar, чтобы визуализировать поведение темы открытого доступа в журналах и другом типе научных документов, которые не заиндексированы в основных международных базах данных, чтобы иметь более широкую и более полную визуализацию изучаемой темы.

Ожидается, что это исследование внесет свой вклад в продвижение знания по рассматриваемой теме, а также что оно поддерживает новые исследования в желании включать такие аспекты, как глобальная идентификация областей, в которых открытый доступ продвинулся больше, типы хранилищ, принятых в разных странах, и вовлеченные учреждения.

ЛИТЕРАТУРА

1. Drott M.C. Open access// Annu. Rev. Inf. Sci. Technol. — 2006. — Vol. 40. — P. 79–109.
2. White S., Creaser C. Trends in scholarly journal prices 2000–2006. — LISU: Loughborough, UK, 2007. — <http://www.lboro.ac.uk/microsites/infosci/lisu/downloads/op37.pdf>.
3. BOAI. Budapest Open Access Initiative.—Open Society Institute: Budapest, Hungary. — <http://www.soros.org/openaccess/read.shtml>
4. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. — Available online: <http://openaccess.mpg.de/>

5. *Bethesda Statement on Open Access Publishing*. — <http://www.earlham.edu/~peters/fos/bethesda.htm>
6. Björk B.-C., Laakso M., Welling P., Paetau P. Anatomy of green open access// J. Assoc. Inf. Sci. Technol. — 2014. — Vol. 65. — P. 237–250.
7. *Graziotin D., Wang X., Abrahamsson P.* A framework for systematic analysis of open access journals and its application in software engineering and information systems// *Scientometrics*. — 2014. — Vol. 101. — P. 1627–1656.
8. *Harnad S., Brody T., Vallières F., Carr L., Hitchcock S., Gingras Y., Oppenheim C., Hajjem C., Hilf E.* The access/impact problem and the green and gold roads to open access: An update// *Ser. Rev.* — 2008. — Vol. 34. — P. 36–40. — <http://eprints.ecs.soton.ac.uk/15852/>.
9. *Laakso M.* Green open access policies of scholarly publishers: A study of what, when, and where self-archiving is allowed// *Scientometrics*. — 2014. — Vol. 99. — P. 475–494.
10. *Rizor S.L., Holley R.P.* Open access goals revisited: How green and gold open access are meeting (or not) their original goals// *J. Sch. Publ.* — 2014. — Vol. 45. — P. 321. — <http://dx.doi.org/10.3138/jsp.45.4.01>
11. *Suber P.* Ensuring open access for publicly funded research// *Br. Med. J.* — 2012. — P. 345. — <http://dx.doi.org/10.1136/bmj.e5184>
12. *Björk B.-C., Welling P., Laakso M., Majlender P., Hedlund T., Guönason G.* Open access to the scientific journal literature: Situation 2009// *PLoS ONE*. — 2010. — Vol. 5, e11273.
13. *Miguel S., Chinchilla-Rodríguez Z., Moya Anegón F.* Open access and Scopus: A new approach to scientific visibility from the standpoint of access// *J. Am. Soc. Inf. Sci. Technol.* — 2011. — Vol. 62. — P. 1130–1145.
14. *Gargouri Y., Larivière V., Gingras Y., Carr L., Harnad S.* Green and gold open access percentages and growth, by discipline. — <http://arxiv.org/abs/1206.3664>
15. *Düzgöl G., Taşkın Z., Tonta Y.* Mapping the intellectual structure of the open access field through co-citation analysis// *Proceedings of the IFLA Satellite Pre-Conference: Open Access to Science Information Trends, Models and Strategies for Libraries, Crete, Greece, 6–8 August 2010.*
16. *Craig I.D., Plume A.M., McVeigh M.E., Pringle J., Amin M.* Do open access articles have greater citation impact? A critical review of the literature// *J. Informetr.* — 2007. — Vol. 1. — P. 239–248.
17. *Codina L.* Scopus: The greatest scientist of the web browser// *Prof. Inf.* — 2005. — Vol. 14. — P. 44–49.
18. *Fingerman S.* Scopus: Profusion and confusion// *Online*. — 2005. — Vol. 29. — P. 36–38.
19. *Jacó P.* As we may search — comparison of major features of the Web of Science, Scopus, and Google Scholar Citation-Based and Citation Enhanced Databases // *Curr. Sci.* — 2005. — Vol. 89. — P. 1537–1547. — <http://www.ias.ac.in/currsci/nov102005/1537.pdf>
20. *Fingerman S.* Web of Science and Scopus: Current features and capabilities// *Issues Sci. Technol. Librariansh.* — 2006. — Vol. 48. — <http://www.istl.org/06-fall/electronic2.html>.
21. *Jacó P.* Comparison of journal impact rankings in the SCImago Journal & Country Rank and the Journal Citation Reports databases// *Online Inf. Rev.* — 2010. — Vol. 34. — P. 642–657.
22. *Bakkalbasi N., Bauer K., Glover J., Wang L.* Three options for citation tracking: Google Scholar, Scopus and Web of Science// *Biomed. Digit. Libr.* — 2006. — Vol. 3, 7.
23. *Falagas M.E., Pitsouni E.I., Malietzis G.A., Pappas G.* Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses// *FASEB J.* — 2008. — Vol. 22. — P. 338–342.
24. *Braun T., Glänzel W., Schubert A.* How balanced is the Science Citation Index's journal coverage? A preliminary overview of macro level statistical data/ *The Web of Knowledge — A Festschrift in Honor of Eugene Garfield, B. Cronin, B., H.B. Atkins (Eds.)*.— *Information Today: Medford, NJ, USA, 2010.* — P. 251–277.
25. *Moya-Anegón F., Chinchilla-Rodríguez Z., Vargas-Quesada B., Corera-Alvarez E., Muñoz-Fernández F.J., González-Molina A., Herrero-Solana V.* Coverage analysis of Scopus: A journal metric approach// *Scientometrics*. — 2007. — Vol. 73. — P. 53–78.
26. *Karino M.E., Felli V.E.* A Enfermagem baseada em evidências: Avanços e inovações em revisões sistemáticas// *Cienc. Cuid. Saúde.* — 2012. — Vol. 11. — P. 11–15.
27. *Hagemann M.* Ten years on, researchers embrace open access. — *Voices*.—<https://www.opensocietyfoundations.org/voices/ten-years-on-researchers-embrace-open-access>
28. *United Nations Educational, Scientific and Cultural Organization.* Global open access portal. The Netherlands. — <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/access-by-region/europe-and-north-america/netherlands/>
29. *Dekker S.* Open Access to publication. — <https://www.government.nl/documents/parliamentary-documents/2014/01/21/open-access-to-publications>
30. *Weitzel S.R.* As novas configurações do Acesso Aberto: Desafios e propostas *Rev. Eletron. Comun. Inf. Inov. Saúde.* — 2014. — Vol. 6. — P. 65–75.
31. *Glänzel W.* National characteristics in international scientific co-authorship// *Scientometrics*. — 2001. — Vol. 51. — P. 69–115.
32. *Person O., Glänzel W., Dannell R.* Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies// *Scientometrics*.— 2004. — Vol. 60. — P. 421–432.
33. *Ennas G., di Guardo M.C.* Features of top-rated gold open access journals: An analysis of the scopus database // *J. Informetr.* — 2015. — Vol. 9. — P. 79–89.

Обогащение онтологий с помощью энциклопедического базового знания для индексирования документов*

Лиза ПОШ
(Lisa POSCH)

Институт социальных наук им. Лейбница,
г. Кельн, Германия

Быстро растущее число научных документов, публично доступных в Интернете, создает проблему эффективной организации и индексирования этих документов. Из-за временных затрат и трудоемкого характера классификации и индексирования, проводимых вручную, существует потребность в лучших методах автоматизации данного процесса. Этот тезис предполагает подход, который усиливает энциклопедическое базовое знание для обогащения онтологий определенной области с помощью текстовой и структурной информации о семантической близости понятий онтологий. Предлагаемый подход стремится изучить эту информацию для улучшения как методов классификации и индексирования документов на основе онтологий, так и методов на основе контролируемого машинного обучения.

ВВЕДЕНИЕ

Число научных публикаций, доступных в Интернете, быстро растет. Без эффективных методов классификации и индексирования документов становится все более затратным по времени и трудным для ученых находить релевантные публикации. Традиционно научные учреждения осуществляли задачу облегчения поиска релевантной литературы путем индексирования и классификации новых публикаций вручную с целью поддержания идеально полной базы данных определенной области. Однако эта задача становится все более трудной для выполнения, поскольку индексирование вручную является затратным по времени, трудоемким и дорогим.

В последние десятилетия ученые различных областей стремятся решать эту проблему путем развития широкого спектра методов автоматической классификации и индексирования текстов. Большинство этих методов основано на алгоритмах машинного обучения или на алгоритмах, которые используют онтологии в качестве базового знания. Хотя существующие подходы позволяют быструю классификацию и индексирование большого числа документов, качество их результатов не сравнимо с эффективностью экспертов-индексаторов. Следовательно, это является постоянной проблемой для улучшения методов автоматической классификации и индексирования.

Главной целью этого тезиса является создание улучшенного подхода к автоматической классификации и предметному индексированию документов при опоре на существующие методы. Предлагаемый подход усиливает энциклопедическое базовое знание для обогащения существующих онтологий определенной области и систем классификации с помощью дополнительной текстовой и структурной информации о семантической близости понятий онтологий. В частности, планируется изучить, является ли это энциклопедическое базовое знание полезным для улучшения результатов методов классификации на основе онтологий и индексирования, а также методов на основе машинного обучения.

Статья имеет следующую структуру. Раздел «Связанная с исследованием работа» дает краткий обзор исследования, относящегося к предметному индексированию и классификации. Затем формулируются вопросы и гипотезы исследования. Далее обсуждается предлагаемый подход и ниже описывается процесс оценки. В разделе «Массивы данных» представляются массивы данных, которые планируется применить в отношении моего подхода; следующий раздел отражает предварительные эксперименты, выполненные на одном из этих массивов данных, и, наконец, подводится итог проделанной работы.

СВЯЗАННАЯ С ИССЛЕДОВАНИЕМ РАБОТА

Большинство подходов к классификации и индексированию документов основано либо на алгоритмах машинного обучения, либо на методах, использующих онтологии в качестве базового знания. Данный раздел кратко суммирует основные техники, которые применялись для изучения проблемы автоматически классифицируемых и индексированных документов.

* Перевод Posch L. Enriching ontologies with encyclopedic background knowledge for document indexing.—
<http://www.arxiv.org/pdf/1603.06494v1.pdf>

Предметное индексирование касается присвоения тематических ключевых слов документам (как правило, из контролируемого словаря, такого как тезаурус), тогда как *классификация документов* приписывает документ к одной или более семантической категории. Разница между этими задачами, тем не менее, незначительна, так как целью обеих является создание соответствующих связей между документами и семантическими объектами [10]. *Семантическое аннотирование* относится к прикреплению метаданных к ресурсам, обычно в контексте семантической сети [12]. Поэтому и предметное индексирование, и классификация документов могут рассматриваться в качестве формы семантического аннотирования документов.

Методы на основе машинного обучения: методы контролируемого и неконтролируемого машинного обучения применяются к классификации документов и индексированию. Популярным методом представления документов в контролируемом обучении является подход *набора слов*, который представляет документы с помощью содержащихся в них слов независимо от порядка, в котором они встречаются. Вместо единичных слов также могут использоваться последовательности слов (*n-граммы*) для представления документа. Слова или *n-граммы* могут быть взвешены различными схемами, такими как частота терминов или TF-IDF (TF - term frequency, IDF - inverse document frequency) [14]. Так как метод TF-IDF не способен охватить семантические структуры в документах, то такие методы как *латентно-семантический анализ (LSA)* [5] и *вероятностный LSA (pLSA)* [8] стремятся преодолеть эту слабость. Также недавно было усилено энциклопедическое базовое знание для представления документов. Например, *явный семантический анализ (ESL)* [6] представляет значение документов как взвешенные векторы понятий на основе Wikipedia. Будучи первоначально предназначенным для вычисления семантического сходства, он также успешно применяется к классификации текстов (например, [7]).

Одним из широко используемых методов неконтролируемой классификации текстов и индексирования являются *тематические модели*. Тематические модели – это статистические модели, стремящиеся открыть латентные темы в документах. Самой простой из них является *латентное размещение Дирихле (LDA)*, которая была введена Блеем и соавторами [3]. Контролируемая версия LDA, *sLDA*, была позднее представлена Блеем и МакОлиффом [2]. Другая контролируемая тематическая модель, *Маркированное LDA*, была введена Рэймиджем и соавторами [13]. Маркированное LDA ограничивает латентные темы, которые должны изучаться на названиях документов в обучающем массиве данных. Тематические модели также используются для создания постоянных рубрик для обучающих контролируемых классификаторов [2].

Методы на основе онтологии: онтология является «*формальной, точной спецификацией общей концептуализации*» [15]. Онтологии использовались как базовое знание для семантического аннотирования документов (например, [9, 4]), большей частью в контексте семантической сети. Жонкэ и соавторы [9] представили открытый биомедицинский аннотатор, который является сетевой службой на основе онтологии для аннотирования документов с помощью биомедицинских понятий онтологии. Процесс аннотации включает два основных шага: первый шаг, *признание понятия*, создает прямые аннотации путем установления соответствия текстовых мета-

данных документов понятиям онтологии. Второй шаг – набор прямых аннотаций *расширяется* за счет использования семантических отношений онтологии и использования существующих отображений других онтологий. Есть намерение создать этот подход и расширить его за счет внедрения текстового и структурного энциклопедического базового знания.

ВОПРОСЫ ИССЛЕДОВАНИЯ И ГИПОТЕЗЫ

Целью исследования является изучение способов, в которых энциклопедическое базовое знание в форме текстовой и структурной информации о семантической близости понятий онтологии может быть полезно для улучшения классификации и индексирования документов. В частности, планируется изучить следующие научные вопросы:

1. Как сравнивается эффективность техник автоматического индексирования и классификации, изучающих энциклопедическое базовое знание, с эффективностью техник, не использующих энциклопедическое базовое знание?

2. Как меняется эффективность техник автоматического индексирования и классификации, изучающих энциклопедическое базовое знание, относительно разных стратегий внедрения базового знания?

3. Каким образом энциклопедическое базовое знание может быть полезно для эффективного сочетания наборов ключевых слов и классов, предложенных методами машинного обучения, с теми, которые предложены методами на основе онтологии?

Предполагается, что энциклопедическое базовое знание о семантической близости определенных понятий в онтологии может быть успешно усилено для моделирования более полного представления передаваемых понятий, и что улучшенное представление этих понятий, вероятно, должно способствовать более точной классификации и индексированию документов.

ПРЕДЛАГАЕМЫЙ ПОДХОД

Этот раздел описывает предлагаемый подход для классификации и индексирования документов с использованием энциклопедического базового знания. Данный подход усиливает энциклопедическое базовое знание, не относящееся к определенной области, чтобы обогатить существующие онтологии определенной области (и системы классификации) с помощью дополнительной информации о понятиях, содержащихся в онтологии. Эта дополнительная информация включает энциклопедическую текстовую информацию о понятиях, которые семантически тесно связаны с понятиями, содержащимися в онтологии, а также структурную информацию о характере семантических отношений между понятиями. Считается, что эта информация может быть полезна для автоматической классификации и индексирования документов, способствуя более точному присвоению семантических классов и тематических ключевых слов с помощью методов на основе онтологии, а также контролируемых методов.

Насколько мне известно, это – первая попытка обогатить существующие онтологии областей и системы классификации энциклопедическим базовым знанием, не относящимся к определенной области, в целях улучшения автоматического индексирования и классификации документов. Два главных шага такого подхода более подробно описываются ниже.

Обогащение существующих онтологий областей

Первый шаг этого подхода состоит в обогащении существующей онтологии области энциклопедическим базовым знанием. Этого можно достичь путем первого соотношения понятий, содержащихся в онтологии, с понятиями, содержащимися в энциклопедии, и впоследствии моделированием семантической близости понятий онтологии. Wikipedia содержит привлекательную опцию для использования в качестве энциклопедии выбора, так как она, как часто оказывается, полезна для широкого диапазона применений в областях, таких как обработка естественного языка, информационный поиск и создание онтологий [1]. Данное соотношение может быть получено либо вручную, либо применением методов автоматического отображения.

Что касается моделирования семантической близости понятий онтологии, то необходимо установить, какие энциклопедические объекты лежат в семантической близости понятий онтологии, а также характер семантических отношений с понятиями онтологии и между объектами энциклопедии. Если Wikipedia выбрана в качестве такой энциклопедии, эта задача может быть решена применением онтологий, извлеченных из Wikipedia, таких как Yago [16] или DBpedia [1]. Одной из главных проблем этого шага является адекватное отображение онтологии целиком и соответствующее обращение с понятиями онтологии, которые не соответствуют ни одному понятию энциклопедии (например, за счет связи их с несколькими родственными понятиями энциклопедии).

Использование обогащенной онтологии с помощью существующих методов классификации и индексирования

Вторым шагом этого подхода является изучение того, полезна ли обогащенная онтология для классификации и индексирования документов. В целях выяснения этого планируется интегрировать ее в существующие методы на основе онтологий и методы контролируемого машинного обучения.

Методы на основе онтологий: в подходе, строящемся на методе, использованном открытым биомедицинским аннотатором [9], энциклопедическое базовое знание, вероятно, будет полезно как в признании понятия, так и в идентификации семантически связанных понятий онтологии для расширения набора ключевых слов. Что касается задачи признания понятия, то текстовая информация из энциклопедических объектов в семантической близости понятий онтологии может быть использована для лучшего установления соответствия понятий в тексте (т.е. путем снятия проблемы словарного несоответствия). Относительно расширения набора ключевых слов структурная информация из энциклопедии может обеспечить поддержку узлов и отношений, которые могут оказаться полезными для принятия лучшего решения по вопросу того, какие дополнительные понятия онтологии включать в набор ключевых слов.

Методы контролируемого машинного обучения: текстовая информация энциклопедических статей в семантической близости понятий категории или онтологии, вероятно, способна ослабить проблему разбросанности в обучающих массивах данных путем предоставления дополнительных обучающих примеров. Возможное ограничение пользы этих дополнительных примеров состоит в том, что используемый в документах тип языка может значительно отличаться от языка, используемого в энциклопедии.

Сочетание методов на основе онтологий и контроля: энциклопедическое базовое знание может оказаться полезным для эффективного сочетания ключевых слов, предложенных контролируемым машинным обучением, и ключевых слов, предложенных подходом на основе онтологий. Окончательные наборы ключевых слов могут сочетаться, например, начиная с помощью промежуточного набора, а затем дальнейшего включения ключевых слов из объединения обоих наборов ключевых слов. Какие ключевые слова должны быть выбраны для индексирования зависит от семантического расстояния, подсчитываемого на базе обогащенной онтологии, к ключевым словам в промежуточном наборе.

ОЦЕНКА

Чтобы оценить пользу обогащенной онтологии для различных методов, планируется использовать три метода, которые будут описаны в этом разделе. Каждый будет применен для сравнения методов индексирования и классификации, которые используют обогащенную онтологию, с соответствующими методами, не использующими обогащенную онтологию.

Оценка на основе существующих, созданных вручную наборов ключевых слов: результаты различных моделей будут сравниваться с существующими, вручную определенными ключевыми словами, предоставленными экспертами-индексаторами. Стандартные метрики, такие как точность, полнота и F1-Score могут использоваться для определения эффективности данных моделей.

Сравнение с семантической близостью на уровне экспертов: измерения семантической близости могут применяться для подсчета семантического расстояния между различными наборами предложенных ключевых слов. Будет измеряться семантическое расстояние между наборами ключевых слов, созданных разными экспертами-индексаторами (для одного и того же документа). Это семантическое расстояние затем будет сравниваться с семантическим расстоянием между набором ключевых слов, созданным этой моделью, и наборами ключевых слов, созданных разными экспертами-индексаторами. Выбор измерения семантической близости может быть основан на предыдущей оценке точности результатов различных измерений, проведенных экспертами области.

Оценка экспертами-индексаторами на основе рекомендации: наборы ключевых слов, созданные моделью, будут представлены экспертам-индексаторам вместе с индекслируемым документом. Затем аннотаторы оценят, какие из предложенных ключевых слов, по их мнению, являются соответствующими, какие не подходят, а какие просто отсутствуют. Основываясь на этой оценке, могут быть подсчитаны стандартные метрики, а также семантическое расстояние в наборе ключевых слов после исправления.

МАССИВЫ ДАННЫХ

Планируется применить описываемый подход и оценить его на следующих массивах данных: информационная система по литературе социальных наук (SOLIS), хранилище открытого доступа по социальным наукам (базы данных, содержащие публикации немецких авторов по социальным наукам), немецкий индекс образования (публикации немецких авторов в области образования) и PubMed Central (публикации английских авторов по биомедицине и наукам о жизни). Первым массивом данных, который будет применяться к данно-

му подходу, является SOLIS, собрание метаданных (включая рефераты) примерно 450 тыс. публикаций по социальным наукам, полностью классифицированных вручную и индексированных экспертами-индексаторами в соответствии с системой классификации и тезауруса по социальным наукам.

ПРЕДВАРИТЕЛЬНЫЕ РЕЗУЛЬТАТЫ

Этот раздел кратко описывает контролируемые эксперименты по классификации, которые проводились на массиве данных SOLIS. Представленные в нем результаты должны считаться не предварительными результатами предложенного подхода, а скорее мотивацией того, почему необходим улучшенный подход.

Формулировка эксперимента: были проведены эксперименты с классификацией на подмножестве базы данных SOLIS, которое содержит все документы, опубликованные после 2003 г. Она содержит 144259 документов и 306879 наименований классов (из системы классификации по социальным наукам). На рис. 1 представлено сжатое распределение классов и понятий онтологии, приписанных к документам. Сжатое распределение классов часто является проблемой при примене-

нии методов контролируемой классификации из-за отсутствия обучающих документов в разбросанных классах.

После применения стандартных методов обработки (удаления стоп-слов и основ слов) подсчитано два набора характеристик на текстовой информации документов: *характеристики TF-IDF* и *тематические распределения маркированного LDA*. Определение категорий, приписанных документу, включает мульти-именную задачу классификации, в которой может быть множество правильных классов для каждого документа. Для этой задачи использовалась стратегия *один против остальных*, которая обрабатывает отдельный классификатор для каждого класса и соотносит этот класс со всеми другими классами. Система классификации по социальным наукам организована иерархически, поэтому можно провести классификацию по различным уровням семантической специфики. Три модели классификации, по одной для каждого из трех верхних семантических уровней в классификационной иерархии, были обработаны для обоих наборов характеристик. Поддерживающие векторные машины с линейными ядрами использовались для всех экспериментов с классификацией, и все классификаторы были обработаны на 67% разбиения подмножества и проверены на оставшихся 33%.

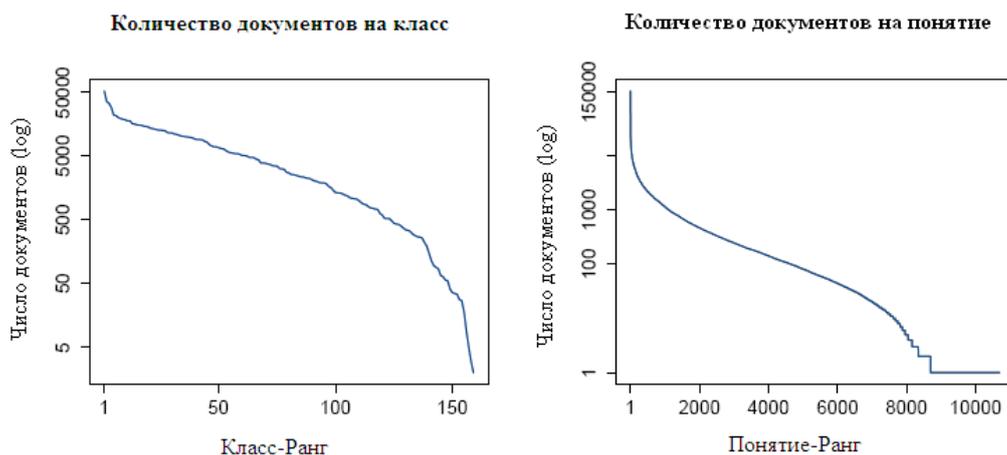


Рис. 1. Распределение документов (логарифмическая шкала) в различных классах из системы классификации для социальных наук и понятий из тезауруса для социальных наук в массиве данных SOLIS.

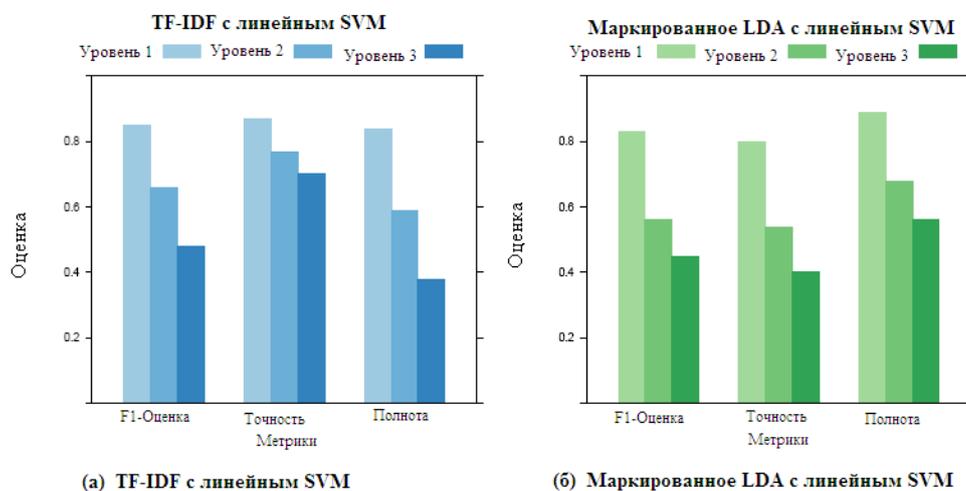


Рис. 2. Результаты контролируемого машинного обучения, нацеленного на три семантически специальных уровня классификационной системы

Результаты и обсуждение: результаты экспериментов представлены на рис. 2. Как правило, характеристики маркированного LDA создают более высокую полноту, а характеристики TF-IDF приводят к более высокой точности. Несмотря на то, что оба набора характеристик достигают принятой оценки F1-Score, когда целью является только первый уровень категорий иерархии, эффективность классификаторов, нацеленных на более семантически точные категории, является неудовлетворительной. Это показывает, что необходим более отработанный подход для эффективной классификации документов по социальным наукам.

ЗАКЛЮЧЕНИЕ

Предметное индексирование неструктурированного текста продолжает составлять проблемную область исследования. Тезис докторской диссертации, представленный в этой статье, базируется на новом подходе к обогащению онтологий и систем классификации с помощью базового знания энциклопедий. Из такого базового знания может быть извлечена текстовая и структурная информация о семантической близости понятий онтологий. Это дополнительное знание при условии более подробного представления понятий, содержащихся в онтологии, вероятно, будет полезным для автоматического индексирования и классификации документов.

Благодарность. Автор выражает признательность профессору, д-ру Маркусу Стромайеру, стороннику представленного в этой статье тезиса.

ЛИТЕРАТУРА

1. Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R., Ives Z. Dbpedia: A nucleus for a web of open data / The semantic web. — Springer, 2007.
2. Blei D.M., McAuliffe J.D. Supervised topic models // Neural Information Processing System. — 2007.
3. Blei D.M., Ng A.Y., Jordan M.I. Latent dirichlet allocation // The Journal of Machine Learning Research. — 2003.
4. De Melo G., Siersdorfer S. Multilingual text classification using ontologies // Advances in Information Retrieval. — Springer. — 2007.

5. Deerwester S.C., Dumais S.T., Landauer T.K., Furnas G.W., Harshman R.A. Indexing by latent semantic analysis // Journal of the American Society for Information Science. — 1990.

6. Gabrilovich E., Markovitch S. Computing semantic relatedness using wikipediabased explicit semantic analysis // Proceedings of the 20th International Joint Conference on Artificial Intelligence. — 2007.

7. Gupta R., Ratinov L.A. Text categorization with knowledge transfer from heterogeneous data sources // Proceedings of the 23rd Conference on Artificial Intelligence. — 2008.

8. Hofmann T. Probabilistic latent semantic indexing // Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. — 1999.

9. Jonquet C., Shab N.H., Musen M.A. The open biomedical annotator // Summit on translational bioinformatics. — 2009.

10. Lancaster F.W. Indexing and abstracting in theory and practice. — University of Illinois Press, 1991.

11. Medehyan O., Milne D., Legg C., Witten I.H. Mining meaning from Wikipedia // International Journal of Human-Computer Studies. — 2009.

12. Oren E., Möller K., Scerri S., Handschub S., Sintek M. What are semantic annotations. — Technical report. — 2006.

13. Ramage D., Hall D., Nallapati R., Manning C.D. Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora // Proceedings of the 2009 Conference on Empirical Methods in Natural Language. — 2009.

14. Salton G., Buckley C. Term-weighting approaches in automatic text retrieval // Information Processing & Management. — 1988.

15. Studer R., Benjamins R., Fensel D. Knowledge engineering: Principles and methods // Data & Knowledge Engineering. — 1998.

16. Suchanek F.M., Kasneci G., Weikum G. Yago: A core of semantic knowledge // Proceedings of the 16th International Conference on World Wide Web. ACM. — 2007.

Приглашаем российских и зарубежных авторов к сотрудничеству
в журнале «Международный форум по информации».
Оригинальные статьи и другие материалы (рецензии, письма)
можно присылать на русском или английском языке
по почтовому адресу, указанному в «Памятке для авторов»
или по электронной почте: mfi@viniti.ru.

Ответственный за выпуск *Л. В. Кобзева*

Компьютерная верстка *М. А. Филимонова*

ИД № 04689 от 28.04.2001 г.

Подписано в печать 02.06.2016 г.

Бумага офсетная. Формат 60x841/8. Гарн. литер. Печать цифровая

Усл. печ. л 6,00 Уч.-изд. л. 6,62 Тираж 33 экз.

Адрес редакции: 125190, Россия, г. Москва, ул. Усиевича, д. 20

Тел. (499) 155-44-95