

СОДЕРЖАНИЕ

Коста М. П. да, Лейте Ф. С. Л. Открытый доступ в мире и Латинской Америке: обзор периода после принятия Будапештской инициативы открытого доступа	3
Джана К., Дутта Б. Представление фасетометрии: новое измерение на основе ключевых слов для расположения научных областей предмета	12
Юджум Э. А., Ратнавелу К., Пратап Г. Лучший из двух миров? Одновременная оценка ученых и их работ	20
Авторский указатель к МФИ, 2016, т. 41	28

**Главный редактор
БИКТИМИРОВ М.Р.**

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

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

Открытый доступ в мире и Латинской Америке: обзор периода после принятия Будапештской инициативы открытого доступа*

Микелли Перейра да КОСТА
(Michelli Pereira da COSTA)

Фернандо Сезар Лима ЛЕЙТЕ
(Fernando César Lima LEITE)

Университет г. Бразилиа, факультет информатики, г. Бразилиа, Бразилиа

В 2012 г. движение за открытый доступ к научной информации отметило десятилетний юбилей своего существования. В целях планирования новых этапов анализировался период, представляющий первую фазу консолидации этого движения. Для обеспечения механизмов и расширения этих дискуссий в статье изучаются исторические аспекты международного и регионального содействия движению открытого доступа. Данный подход был основан на описании основных имевших место во времени событий и инициатив, отраженных в научной литературе по данной теме. Элементы, обсуждаемые в исследовании, были организованы согласно двум параметрам. Первый относится к временной перспективе, отсчитываемой от публикации Будапештской инициативы открытого доступа. Второй – затрагивает краткое обобщение участия Латинской Америки. В качестве результата представляется временной период открытого доступа в мире и в Латинской Америке, охватывающий основные, затронутые исследованием аспекты. Цель исследования достигается за счет предлагаемой систематизации, так как анализируются инициативы открытого доступа в Латинской Америке и определяется то, как они сами влияли на него и подверглись влиянию других регионов мира.

ВВЕДЕНИЕ

Процесс научной коммуникации постоянно меняется. Эти изменения в основном происходят благодаря социальным контекстам и технологическим усовершенствованиям. Согласно Медоузу [1], самые значимые изменения происходили в результате появления журналов в середине XVII в., научных сообществ в XVIII в., становления профессионализма и экспертизы в научно-исследовательских областях знания в XIX в. и возникновения позже электронного мира в XX в.

Наравне с факторами, вызвавшими изменения в предыдущие века, два других сочетающихся элемента активизировали важные перемены в системе научной коммуникации в XXI в. Первый касается неудовлетворенности ученых традиционной моделью научной публикации, ограничивающей доступ к их научным результатам. Второй в свою очередь характеризуется появлением

возможностей расширить и рационализировать коммуникационные процессы, обусловленные новейшими технологиями, уже внедренными в мир научной коммуникации, начиная с последнего столетия. Эти два фактора, а именно неудовлетворенность ученых и технология, стимулировали появление инициатив открытого доступа.

Инициативы открытого доступа относятся к двум стратегиям. Одна, известная как золотая дорога, поощряет научную публикацию в журналах открытого доступа. Другая, именуемая зеленой дорогой, стремится развивать цифровые хранилища открытого доступа, из которых публикации или принятые к публикации статьи распространяются альтернативно. Обе стратегии были усилены созданием открытого доступа.

Это движение началось с заявления ученых и институтов большей частью из северных стран. Однако данная перспектива своевременна для стран социально и исторически исключенных из центра науки. Среди преимуществ открытого доступа самым важным является то, что ученые, чьи институты не в состоянии платить за подписку на коммерческие журналы и базы данных, могут иметь доступ к литературе. Более того, ожидается, что эти инициативы усилят обозримую литературу открытого доступа.

* Перевод Costa M. P. da, Leite F.C. L. Open access in the world and Latin America: A review since Budapest Open Access Initiative.— <http://eprints.rclis.org/28881/1/0103-3786-tinf-28-01-00033.pdf>

Что касается производства, распространения и доступа к научному знанию, то при сравнении с другими регионами Латинская Америка обладает уникальными особенностями, благоприятствующими принятию открытого доступа. Это отмечается многими авторами [2-4].

Чен и Коста [2] подчеркивают важность доступа к науке и технике ради сглаживания проблем развивающихся стран, как в случае Латинской Америки. По мнению авторов, в этих странах заморожено производство науки и техники из-за слабого инвестирования исследований и трудностей доступа к мировой научной литературе по причине высоких цен подписки на основные журналы. Авторы [2] представили опубликованные ВОЗ данные за 2004 г., показывающие, что 56% стран с валовым внутренним продуктом ниже 1 тыс. долл. не подписывались на основные журналы в 1999-2004 гг. Учитывая эти реалии, открытый доступ, по их мнению, был ключевым элементом важных изменений в научной коммуникации, так как он предоставляет более справедливый доступ к научным публикациям и включает участие ученых из развивающихся стран в «глобальной интеллектуальной общине», т.е. глобальном общественном благе.

Альперин и др. [3] в свою очередь обращают внимание на количество журналов открытого доступа, опубликованных в Латинской Америке, в частности в Бразилии. Согласно данным дирекции журналов открытого доступа на октябрь 2013 г., Бразилия является второй страной с самым большим числом журналов открытого доступа [5]. Однако авторы отмечают, что причины, послужившие росту открытого доступа в регионе, отличаются от причин, возникших в северных странах. Это объясняется тем, что журналы являются традиционно доступными в этом регионе, где нет крупных коммерческих издателей, таких как в североамериканских и европейских странах. Таким образом, для Латинской Америки появление информационной технологии и инициатив открытого доступа означает шанс создания необходимой инфраструктуры в целях расширения распространения научного производства, так как она никогда не фокусировалась на прибыли.

Бабини [4], помимо прочего, отмечает развитие институциональных хранилищ в регионе. По его мнению, учреждения, производящие научное знание, университеты и исследовательские институты считают эти системы «адекватным выбором» для управления и обозримости институционального научного производства. Более того, некоторые правительства стран, таких как Аргентина, Бразилия, Перу и Мексика, поддерживают институциональные хранилища.

Открытый доступ вносит вклад в научное развитие посредством оптимизации некоторых процессов и поощряет возможность других регионов участвовать в международной научной коммуникации. Среди новых регионов Латинская Америка демонстрирует сильный интерес к открытому доступу. Поэтому, учитывая обсуждаемые далее особенности и возможности, цель настоящего исследования заключалась в анализе инициатив открытого доступа в регионе и рассмотрении того, как они повлияли и подверглись влиянию инициатив из других регионов мира.

МЕТОДЫ

Мы адаптировали библиографический метод исследования, чтобы достичь цели, предложенной Лимой и Миотой [6]. По мнению авторов, этот тип исследования «подразумевает структурированный набор процедур для поиска решений относительно объекта изучения». Для

проведения исследования авторы предложили следующие четыре параметра анализа литературы: тематический, лингвистический, хронологический и параметр источников.

Для тематического и лингвистического параметров отбирались исследования, написанные на английском, португальском и испанском языках, затрагивающие в основном исторические элементы открытого доступа и представляющие очевидность участия Латинской Америки. Базы данных Scopus и Google Scholar использовались как источник сбора данных. Мы выбрали эти две базы данных, так как они широко применяются в качестве новых источников библиографических метрик, что обсуждалось Мехо и Яном [7]. По мнению этих авторов, обе базы данных считались альтернативными и дополняющими по отношению к традиционным и гибким средствам, таким как Web of Science.

Мы провели поиск ключевых слов для терминов «история»+«открытый доступ»+«Латинская Америка» на португальском, английском и испанском языках в двух базах данных. В базу данных Google Scholar был добавлен параметр тип файла: pdf, чтобы система восстановления отдавала предпочтение файлам PDF. Применение поисковых стратегий выдало 4914 документов в Google Scholar и 56 – в Scopus.

Предварительное чтение отобранных текстов позволило идентификацию ссылок других релевантных авторов, чьи исследования были включены в анализ. Исследования, рассматривавшиеся в качестве источника данных, включали следующие документы: Харнад [8, 9], Оальджко [10], Сарменто и др. [11], Аски [12], Чиваллеро [13], Бейли [14], Альперин и др. [3], Мюррей-Раст [15], Баррос [16], Будапештская инициатива открытого доступа [17], Мишра [18], Сьюбер [19].

На основе этих исследований мы подготовили историческое описание основных релевантных фактов, способствовавших установлению открытого доступа в мире, проливающее свет на действия в Латинской Америке. Глобальная и региональная параллель, иллюстрирующая период открытого доступа в мире и в Латинской Америке, показана в конце статьи. Затем приводятся результаты.

ИСТОРИЧЕСКИЕ ПЕРСПЕКТИВЫ ОТКРЫТОГО ДОСТУПА

Открытый доступ подразумевает свободную и неограниченную доступность полных научных публикаций в Интернете. Его консолидация в мире поощрялась международными усилиями сторонников открытого доступа. Выражение «открытый доступ» впервые было использовано для этой цели Будапештской инициативой открытого доступа [20], которая поэтому будет использоваться в качестве подхода к обсуждению открытого доступа в данной статье. По этой причине историческая перспектива открытого доступа поделена на три периода: до Будапештской инициативы открытого доступа, в период Будапештской инициативы открытого доступа и после Будапештской инициативы открытого доступа.

До Будапештской инициативы открытого доступа. Хотя Будапештская инициатива открытого доступа считается знаковым документом для открытого доступа, она не была первой инициативой для обсуждения основной идеи открытого доступа. Сарменто и др. [11] придают большое значение трем предыдущим документам, появившимся в результате ряда встреч во время Всемирной конференции по науке XX в. Первый, «Декларация Санто-Доминго» [21], отмечает, что наука должна быть доступна каждому и поэтому инфраструктуру возможно

стей информационных технологий и распространения науки следует усилить, чтобы она стала доступной обществу. Второй, «Декларация по науке и использованию научного знания», и третий, «Повестка дня по науке», документы фокусируют внимание на необходимости обмена научной информацией, особенно создаваемой совместно с общественными фондами, признавая, что доступ к этой информации ценен для прогресса самой науки. По мнению Сарmento и др. [11, с. 3], эти документы релевантны при рассмотрении контекста, в котором это движение развивается, поскольку они обращают внимание на создание «нового социального контракта для науки, проливающего свет на компоненты науки и общества, и инновации для обеспечения развития конкретных инициатив международного научного сотрудничества».

Ориентируясь на предшествующие, до 1999, годы, Чиваллеро [13] также отмечает другие инициативы. Автор рассматривает хранилище ArXiv как первую релевантную попытку, разработанную в 1991 г. По мнению Рамло [22], это хранилище было разработано Полем Гинспаргом под названием LANL Preprint Archive, относящимся к институту – Лос-Аламосская национальная лаборатория (Los Alamos National Laboratory, LANL), которая его разместила. Когда Гинспарг покинул лабораторию, хранилище принял Корнеллский университет, содержащий литературу открытого доступа по физике и смежным наукам.

Вслед за развитием ArXiv Чиваллеро [13] отмечает встречу, состоявшуюся в 1992 г. и организованную Ассоциацией научных библиотек при инициативе Коалиции академических ресурсов и научных изданий, обсуждавшей литературу в электронном формате. По этому случаю рассматривались новый контекст обработки и использование научных публикаций в электронных медийных средствах. Учитывая развитие систем открытых файлов и новый сценарий, Стивен Харнад опубликовал в 1994 г. исследование, озаглавленное «Сенсационное предложение» («The subversive proposal»), обобщающее идею самоархивирования. Развитие этой идеи привело к тому, что открытый доступ позже назвал стратегией зеленой дороги.

В 1999 г., в том же году, когда были опубликованы декларации по науке, в США была подписана Конвенция Санта-Фе. Цель, по мнению Лагозе и Ван ден Сомпеля [23], состояла в том, чтобы обсудить возможности использования сети научно-информационных систем для придания большего динамизма и скорости в коммуникации. Итогом этой встречи стала организация инициативы открытого доступа, суть которой состоит в использовании протокола инициативы открытых архивов для протокола сбора метаданных. Этот протокол коммуникации позволяет системам обмениваться метаданными (данными, описывающими информационные источники) без посреднических процессов. Такая инициатива была важна для развития информационных систем открытого доступа, так как она установила стандарты и протокол для управления и обеспечения взаимодействия между системами, одного из строительных блоков для продвижения и обзорности научной литературы в среде открытого доступа.

В период между 2000 и 2001 гг. с целью обсуждения вопросов науки и доступа к её результатам проводились и другие встречи при участии ученых и информационных специалистов. В результате этих событий были опубликованы обсуждаемые нами ниже декларации. Открытое письмо публичной научной библиотеки (PLOS) и документ, названный «Провозглашением неза-

висимости», опубликованный Коалицией академических ресурсов и научных изданий, были самыми выдающимися, так как касались важных институтов, поощряющих открытый доступ.

Декларация PLOS является открытым письмом нескольких академических институтов в области биомедицинских наук. Она излагает свою поддержку развития публичной онлайн библиотеки, обеспечивающей полные тексты научных публикаций. В итоге, подписавшие этот документ обязались публиковаться только в журналах, разрешающих свободное и неограниченное повторное распространение публикаций в системе PubMed Central (PMC) или иных подобных информационных системах. PMC является хранилищем, созданным в 1999 г. национальными институтами здравоохранения (США) для хранения и распространения рукописей и опубликованных исследований в области медицинских наук [24]. Это служит еще одним примером успешной инициативы открытого доступа, появление которой предшествовало Будапештской инициативе открытого доступа.

Документ «Провозглашение независимости», в свою очередь, является письмом, адресованным ученым, но подписанным только одним из них – Майклом Розенцвейгом. Хотя письмо подписано одним ученым, оно было опубликовано Коалицией академических ресурсов и научных изданий и Сетью научных библиотек трех университетов, главными международными организациями научного сообщества. Одним из ключевых аргументов является обсуждение доступа к научной информации, связанной с кризисом сериальных изданий. Автор прежде всего обращает внимание на то, что их исследование продается коммерческими издателями с планами по увеличению прибыли, и утверждает, что подписки слишком дороги для библиотек. Согласно рисунку, предоставляющему данные за 1986-1998 гг., журнальная подписка в области эволюции экологии увеличивалась примерно на 800 долл. в год, достигнув 75% профицитной маржи. Опираясь на этот аргумент, автор приводит пример развивающегося «независимого» журнала в области, созданного в качестве альтернативы коммерческой модели.

В Латинской Америке инициативы, предшествующие Будапештской, разделяли беспокойства относительно ограниченного доступа к научной информации и интерес в использовании Интернета для ускорения и демократизации распространения информации. Мы осветим три из этих инициатив. Первая в хронологическом порядке – «Декларация Сан-Хосе, касающаяся виртуальной библиотеки в области здравоохранения», появилась в 1998 г. По этому документу центры Латинской Америки и стран Карибского бассейна по информации в области медицинских наук были обязаны разработать Виртуальную библиотеку здравоохранения для усиления доступа к информации в области. Виртуальная библиотека здравоохранения, учрежденная в том же году, что и декларация, в соответствии с ее собственной концепцией является научной и технической информационной системой для стимуляции создания и использования информационных источников в электронном формате через Интернет с целью продвижения прямого и универсального доступа.

В 2001 г. «Гаванская декларация по справедливому доступу к медицинской информации» снова отражает ответственность государства относительно доступа к информации. В этой декларации участники встречи в рамках Виртуальной библиотеки в области здравоохранения классифицировали научную информацию по данной тематике как глобальное общественное благо и подчеркнули необходимость национальных и междуна-

родных политик для обеспечения доступа и распространения. Таким образом она повторно подтвердила свое обязательство в отношении развития Виртуальной библиотеки здравоохранения.

Помимо Виртуальной библиотеки здравоохранения, была начата в 1997 г. разработка научной электронной онлайн библиотеки (SCiELO). SCiELO индексирует журналы, финансируемые государственным фондом поддержки исследований. Индексируемые журналы состоят из полных, свободно доступных текстов жур-

нальных собраний. Согласно данным SCiELO за 2013 г., система содержала массивы журналов из 11 стран, восемь из которых были странами Латинской Америки.

Все эти события, документы и системы важны для первоначальной консолидации открытого доступа. Поэтому число инициатив и своевременно приводимых аргументов для обсуждения альтернатив по обеспечению широкого доступа к научным публикациям были представлены в ходе подготовки Будапештской инициативы открытого доступа.

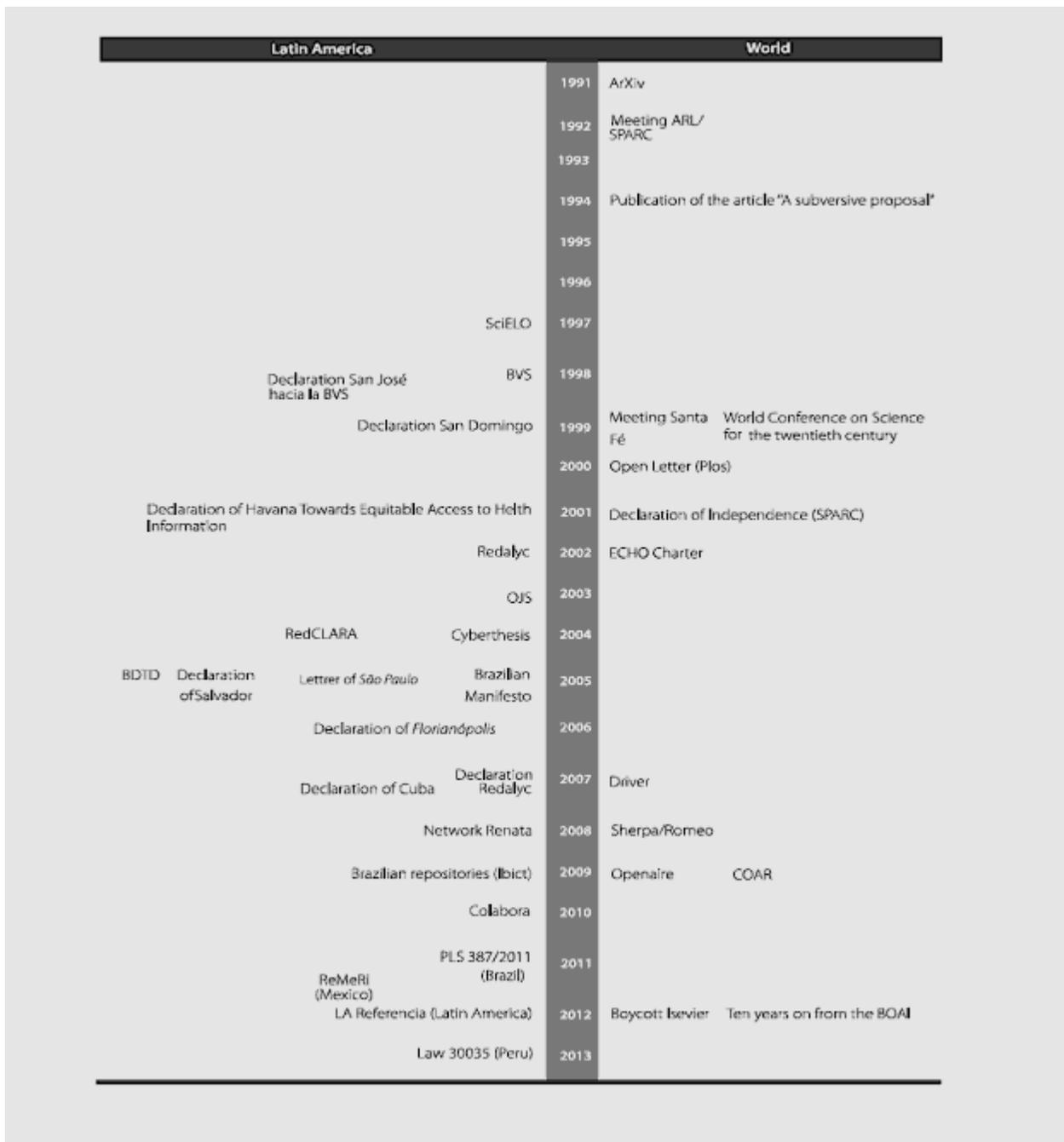


Рис. Временной период развития открытого доступа в мире и Латинской Америке

Источник: Costa M. P. da, Leite F.C. L. (2014).

Примечание: ArXiv: Служба электронных изданий в областях физики, математики, вычислительной техники, количественной биологии, количественных финансов и статистики; BOAI: Будапештская инициатива открытого доступа; BDTD: Бразильская цифровая библиотека тезисов и диссертаций; BVS: Виртуальная библиотека в области здравоохранения; COAR: Конфедерация хранилищ открытого доступа; ECHO: Культурное наследие в режиме онлайн; OJS: Система открытых журналов; PL: Проект закона (Бразилия); PLS: Проект закона и Сената (Бразилия); SNRD: Национальная система цифровых хранилищ (Аргентина); SPARC: Коалиция академических ресурсов и научных изданий.

Будапештская инициатива открытого доступа. Основываясь на упомянутых выше дискуссиях и инициативах, в 2001 г. в Будапеште состоялась встреча, и документ относительно Будапештской инициативы открытого доступа был опубликован в феврале 2002 г. [20]. Документ затрагивает проблему научной коммуникации в современном контексте и считает, что процессы традиционной коммуникации через журналы должны подвергнуться влиянию новых технологий и Интернета. В этом контексте позитивное влияние понимается как электронное распространение полных текстов из рецензируемых журналов через Интернет и без ограничений доступа. При обсуждении понятия открытого доступа, исходя из первого опыта, утверждается, что эта модель возможна. Отдельно примеры опытов не приводятся, но отмечается, что различные инициативы все еще предоставляли «чрезвычайную власть» читателям, так как они могли находить и использовать релевантную литературу. Более того, согласно Декларации, исследования, доступные в системах открытого доступа, и авторы получают большую обзорность, читательскую аудиторию и влияние.

Преимущества, излагаемые в Будапештской инициативе открытого доступа, периодически используются для демонстрации выгод открытого доступа. В документе, доступном в Интернете, преимущества (обзорность, читательская аудитория и влияние) гиперсвязаны с другими текстами, обсуждающими эту проблему. Первым текстом, получившим гиперссылку в Будапештской инициативе открытого доступа, была статья Харнада [9], изучающая вопросы, относящиеся к доступу, влиянию и оценке публикаций. Автор утверждает, что цель ученых – публикация статей ради получения влияния, а не их продажа, поскольку они (ученые) «вознаграждаются» университетами и финансирующими организациями, дающими оценку влияния того, что они производят. Таким образом, когда результаты их исследований становятся публично доступными, это повышает импакт-фактор, который представляет интерес для ученых и институтов.

Кроме преимуществ открытого доступа, Харнад [9] также показывает, что онлайн-самоархивирование рецензируемого произведения в институциональных системах является стратегическим. Однако автор обращает внимание на то, что самоархивирование большей частью признано институтами успешным. Харнад открывает дискуссию о необходимости институциональных политик открытого доступа, являющихся формальными инструментами, отражающими позицию института относительно обращения его научной продукции.

Вторая и третья гиперссылки на Будапештскую инициативу открытого доступа [20] приводят к ошибке в указании страницы, которая не позволяет идентификацию текстов. Попытки получения доступа к гиперссылке предпринимались в июле и августе 2013 г. Термин «влияние» является четвертой гиперссылкой на статью Одадьжко [10], изучающую эволюцию научной коммуникации. Направление в сторону использования онлайн-контента является одним из исследуемых автором факторов, преследующим цель продемонстрировать то, какое влияние могут иметь новые динамики коммуникации. Согласно представленным в статье данным, использование этих контентов постепенно растет. В качестве примера такой тенденции автор приводит появление электронных тезисов и диссертаций, а также опыт SCiELO, обзорность которой выросла с 5 тыс. стр. в 1999 г. до 64 тыс. стр. в 2000 г. Несмотря на высокий рост темпа доступа к этой системе, автор утверждает, что она не дает одинаковое количество доступа при

сравнении с печатными публикациями, даже если и близка к ней.

После обсуждения преимуществ открытого доступа Будапештская инициатива [20] приглашает отдельных лиц и институты к сотрудничеству с помощью предоставления открытого доступа к академической литературе. Поэтому декларация определяет, какие лицензии следует дать публикациям в этом контексте, учитывая, что единственное ограничение, которое необходимо предоставить авторам, – это контроль целостности их работы и права быть должным образом цитируемыми.

Для достижения своих целей Будапештская инициатива открытого доступа [20] предложила две стратегии: самоархивирование и развитие журналов открытого доступа. Самоархивирование подразумевает, что авторы могут загружать копию своих отрецензированных публикаций в хранилища открытого доступа, которые должны быть в согласии со стандартами инициативы открытого доступа. Позже это предложение стало известным как стратегия зеленой дороги. Вторая стратегия предполагает изменение в бизнес-модели существующих журналов и развитии журналов открытого доступа, которой следует искать другие формы дохода вместо взимания платы за доступ к своему контенту. Применение этих руководств по отношению к журналам в последующем стало известно как золотая дорога к открытому доступу.

Будапештская инициатива [20] поднимает проблему открытого доступа к научной коммуникации и обращает внимание на стратегии совместных действий. Для придания законной силы этому документу декларация была подписана 17 университетскими учеными, научно-исследовательскими центрами и издателями в Северной Америке и Европе и поэтому, по мнению Сарmento и др. [11], научное сообщество приняло на себя обязательство в отношении открытого доступа.

Более того, Будапештская инициатива открытого доступа в соответствии с новой редакцией, опубликованной 10 годами позже [17], была первой международной акцией, цель которой получить значительное финансирование на развитие стратегий. Однако документ не оговаривает, какое финансирование было получено и кто его предоставил.

Все эти факторы показывают, насколько важным элементом для обсуждений открытого доступа была Будапештская инициатива открытого доступа. По мнению Сарmento и др. [11, с. 4], «поскольку документ представляет некоторые принципы, стратегии и условия выполнения обязательств, он стал самым важным документом в движении открытого доступа». Кроме того, Будапештская инициатива открытого доступа широко цитируется в следующих ей документах, а ее руководства учитываются при развитии систем открытого доступа.

После Будапештской инициативы открытого доступа. В следующем году после публикации Будапештской инициативы открытого доступа в 2003 г. были опубликованы, как отметили Сарmento и др. [11], две другие декларации с особым вниманием по отношению к открытому доступу. Значение этих двух деклараций вместе с важностью Будапештской инициативы открытого доступа также демонстрируется другими авторами, которые обсуждают этот вопрос, такими как Аски [12], Бейли [14], Мюррей-Раст [15], Мишра [18] и Сьюбер [19].

Одним из этих документов стало Бетесдское заявление по издательству открытого доступа [25]. Оно было подготовлено во время собрания Медицинского института Говарда Хагса в шт. Мэриленд, США. Цель – вызвать дебаты по открытому доступу в научном сообще-

стве по биомедицине и попытаться достичь соглашения по этому вопросу с помощью привлечения всех участников процесса научной коммуникации. Помимо предложенных действий, заявление оговаривает два условия того, чтобы публикация считалась открытым доступом. Первым условием является разрешение автора, дающее пользователям право на копирование, использование, распространение, передачу, изображение и создание и распространение производных работ. Второе - непосредственная запись в файл копии публикации в цифровое онлайн хранилище научного института или организации таким образом, чтобы она согласовывалась со стандартами интероперабельности и хранения. В рамках этих условий заявление рассматривает хранилище РМС в качестве хранилища в области биомедицинских наук.

Вторым документом является Берлинская декларация по открытому доступу к знанию в науках и гуманитарных сферах [26], подписанная исследовательскими организациями и учеными разных стран. Важный аспект Берлинской декларации состоит в явном расширении выгод и потребности в открытом доступе в областях знания, связанных с социальными науками и гуманитарной сферой. Заявление требует согласования с Будапештской инициативой открытого доступа, ЕСНО Charter и Бетесдской декларацией. Цель Берлинской декларации – вдохновить ученых на публикацию в открытом доступе, поддержку перехода к новой парадигме научной коммуникации, поощрять публикации открытого доступа и быть на их стороне в признании академических процессов оценки. Последнее требование подчеркивает важность финансирующих организаций в консолидации действий открытого доступа.

Кроме упомянутых выше, были выпущены другие декларации, усиливающие движение открытого доступа. Эти новые документы следовали модели предыдущих деклараций: они были документами ученых, участвовавших в событиях академической жизни, и поэтому получили названия городов или регионов, где происходили эти мероприятия: Декларация Вальпарайсо по совершенствованию научной коммуникации в электронной среде [27], Принципы округа Вашингтон относительно свободного доступа к науке [28], Брюссельская декларация по открытому доступу [29], Хорватская декларация открытого доступа [30]. Другие документы относились к обязательствам институтов в этом вопросе: Декларация шотландских университетов по открытому доступу [31], Принципы и руководства по вопросу доступа к исследовательским данным из публичного финансирования (Организация экономического сотрудничества и развития, [32]), Заявления ИФЛА по открытому доступу – разъяснение позиции и стратегии ИФЛА (Международная федерация библиотечных ассоциаций и институтов, [33]), Международная ассоциация COMMUNIA по вопросу позиции в общественной сфере в отношении политики открытого доступа ЕС Горизонт 2020 (Международная ассоциация COMMUNIA в общественной сфере, [34]), Рекомендация по открытому доступу и онлайн контенту в базе данных по наукам здравоохранения в странах Латинской Америки и Карибского бассейна [35]. Документы большей частью представляли открытый доступ и давали объяснение причин, почему важно развивать действия, поощряющие широкий доступ к научной информации, все они сводились к обязательствам и стратегиям по достижению цели.

В 2012 г. Будапештская инициатива открытого доступа отметила 10 лет существования и опубликовала

новый документ «Десять лет после Будапештской инициативы открытого доступа: установка на открытость по умолчанию» [17]. В этой редакции подтверждены стратегии, предложенные Будапештской инициативой открытого доступа: зеленая дорога к архивам открытого доступа и золотая дорога к журналам открытого доступа. Дополнительно представлены новые рекомендации на следующие десять лет по использованию политик открытого доступа институтами, производящими научное знание, лицензированию контента открытого доступа, защите, координации инициатив и продвижению систем открытого доступа.

В том же году, когда отмечалось 10 лет Будапештской инициативы открытого доступа, по всему миру имело место новое масштабное движение, требовавшее открытого доступа к научной литературе. Оно началось с предложения ученого Тимати Говерса бойкотировать издательство Elsevier, которое к концу 2014 г. должно иметь в целом 15 тыс. подписчиков. Целью бойкота, по мнению Барроса [16], был принятый благодаря оказанной издателем поддержке американский законопроект, пересмотревший политику доступа национальных институтов здравоохранения, неудачно пытающихся продвигать открытый доступ к публикациям, сообщаящим результаты исследований, финансируемых государственным фондами. В результате бойкота, по мнению автора, десять издателей вышли из Elsevier, поддержка законопроекта была отозвана, а новый законопроект, поощряющий открытый доступ, набирает обороты в США.

ОТКРЫТЫЙ ДОСТУП В ЛАТИНСКОЙ АМЕРИКЕ

В Латинской Америке, как объяснялось ранее, дискуссии и системы развивались вокруг исследуемого вопроса задолго до Будапештской инициативы открытого доступа [20]. В период после Будапештской инициативы открытого доступа самой сильной инициативой в его пользу был Бразильский манифест в поддержку открытого доступа к научной информации [36]. Манифест выпущен Бразильским институтом информации по науке и технике в 2005 г. В нем использовались условия Берлинской декларации, относящиеся к способам публикации и условиям для открытого доступа. Манифест освещает потребность в участии академических институтов, исследователей, финансирующих организаций, коммерческих и некоммерческих издателей. После упоминания действий, которые каждый из вовлеченных участников должен принять во внимание, в конце манифест указывает на важность создания национальной политики для обеспечения выполнения этих действий.

В 2005 г. была опубликована «Сальвадорская декларация по открытому доступу: перспектива развивающихся стран» [37]. Она подготовлена участниками Международного семинара по открытому доступу. Документ подчеркивает ожидание роста открытого доступа в развивающихся странах и, как следствие, возможность ученых получить доступ к литературе и науке. Далее представляются требования подписавшихся относительно научных политик.

Позже в 2005 г. было опубликовано «Письмо шт. Сан-Паулу», являющееся проявлением поддержки открытого доступа. Письмо содержит аргументы в пользу предложения открытого доступа, среди которых выражена идея, что доступ к научной литературе необходим для научного и социального прогресса. В целях гарантии доступа на рассмотрение предложены шестнадцать

рекомендаций, а также стратегии, выдвинутые Будапештской инициативой открытого доступа и ее новой редакцией.

Декларация Флорианополиса была предварительно составлена бразильскими учеными в области психологии в 2006 г. с целью выразить поддержку открытому доступу к рецензируемой литературе. Этот документ отсылает к понятию открытого доступа, изложенного в Бетеседе [15], также предлагаются некоторые рекомендации по достижению обозначенных целей.

С той же самой задачей было выпущено много других документов в Латинской Америке. Мы осветим Кубинскую декларацию о пользе открытого доступа [38]. Этот документ ссылается на термины, используемые Бюро по улучшению деловой практики, Сальвадором и Флорианополисом. Декларация приглашает ученых делать свои работы доступными в институциональных хранилищах и цифровых библиотеках открытого доступа и освещает Виртуальную библиотеку Союза латиноамериканских психологических организаций.

Законопроект № 1120 от 2007 г. [39] был представлен в бразильский конгресс. Он предлагает создание национального закона для обязательного развития информационного поиска учреждениями высшего образования и государственными научно-исследовательскими центрами. Цель законопроекта – развитие информационного поиска и обязательная публикация копий проведенных учеными исследований, получивших государственное финансирование. Законопроект получил одобрение комитета с некоторыми поправками, но в 2011 г. был отправлен в архив, так как находился в конгрессе четыре года и не рассматривался на повестке дня [40]. В том же году он был заменен сенатским законопроектом № 387 [39]. Его основное содержание идентично содержанию законопроекта 2007 г., но имеет ряд уточнений, таких как тип депонируемых документов и срок между публикацией и депонированием. Законопроект еще не одобрен, но согласно portalу законодательной деятельности сената, он рассматривается в повестке дня различных комитетов.

Подобная инициатива была учреждена в Аргентине. В 2011 г. по решению министерств была создана Национальная система цифровых хранилищ. Ее цель – разработать взаимодействующую сеть цифровых хранилищ с помощью учреждения политик, стандартов и общих протоколов [41]. В том же году был подготовлен законопроект, одобренный в 2012 г., чтобы гарантировать развитие информационного поиска в институтах, получающих государственное финансирование, и депонирование копий как рецензированных опубликованных, так и не опубликованных работ. Новизной этого законопроекта по отношению к бразильскому является условие обращения первичных данных исследования.

В Мексике также в 2011 г. была создана сеть Red Mexicana de Repositories Institucionales (ReMeRI - Мексиканская сеть институциональных хранилищ), имеющая интегрированную поисковую систему по информационному поиску и предлагающая обучающие семинары и разработку сервисных программ. Что касается аргентинской системы, то здесь требуются некоторые стандарты для подключения к Мексиканской сети, которые диагностируются посредством инструмента надежности [42].

В 2012 г. была основана Федеральная латиноамериканская сеть институциональных хранилищ по научной документации в Латинской Америке (Red Federada Latinoamericana de Repositorios Institucionales de Documentación Científica en América Latina, LA Referencia) при поддержке RedCLARA (сеть сотрудничества лати-

ноамериканских перспективных сетей – Red Cooperación Latino Americana de Redes Avanzadas) и спонсорстве регионального фонда общественных благ Межамериканского банка развития. Ее цель – содействовать созданию региональной стратегии информационного поиска. В связи с этим были подняты вопросы, касающиеся соглашений, региональных политик и определения стандартов [4]. В мае этого года Латинская Америка через LA Referencia присоединилась к Конфедерации хранилищ открытого доступа, как представлено на их сайте [43]. Конфедерация является ассоциацией академических институтов и исследований из европейских, азиатских, североамериканских и латиноамериканских стран, нацеленной на то, чтобы усилить глобальные сети хранилищ открытого доступа.

В следующем 2013 г. законопроект № 30035 [44] по открытому доступу был одобрен в Перу. Он регулирует работу национального цифрового хранилища по науке, технике и инновациям в открытом доступе. Данное хранилище характеризуется как «центральный сайт», где хранится цифровая информация, касающаяся производства науки, технологии и инновации в стране. Его использование адресовано исследователям, принадлежащим к государственным организациям, исследователям, которые получают государственное финансирование, а также исследователям, желающим сделать свою продукцию свободной для открытого доступа.

Национальные законы, такие как эти, безусловно являются мощными союзниками по обеспечению появления и/или консолидации механизмов, продвигающих открытый доступ. В случае законопроектов в Латинской Америке была предпочтительной стратегия зеленой дороги. Некоторые ученые - сторонники открытого доступа, такие как Сьюбер [45], утверждают, что институциональная эффективность лучше, если используется стратегия зеленой дороги. Так как право ученых публиковаться там, где они хотят, гарантируется обязательством в дальнейшем депонировать копию в хранилищах открытого доступа по соглашению с издательством.

В Латинской Америке эффективные сети были разработаны для внедрения информационного поиска, включающего международные стандарты взаимодействия и сохранения. Развитие работ в сетях хранилищ может облегчать обмен ресурсами, разрешая более высокое качество обмена данными и обеспечивая большую прозрачность научного производства. Сетями информационного поиска, расположенными в регионе, являлись: Comunnidade Colabora, La Reference, SNRD и ReMeRI.

Эффективность Латинской Америки в региональных сетях может объясняться особенностями региона, которые отличаются от стран в Северной Америке и Европе. При обсуждении этих особенностей Альперин и др. [3] представляют данные, показывающие рост в инфраструктуре региона по сравнению с другими регионами. Однако рост не являлся достаточным, чтобы обеспечить хорошую инфраструктуру для информационных систем, так как Интернет до 2008 г. был доступен только 22,1% населения.

Кроме вопроса относительно инфраструктуры в регионе, препятствующей строительству более сильных механизмов распространения произведенной здесь науки, отмечается, что историческая и социальная формация также подвержена проблемам. Эти элементы должны учитываться при рассмотрении произведенной в регионе науки, так как они являются структурой системы. По мнению Суньити [46], в развивающихся странах и в частности в Латинской Америке реальность науки и

техники тесно связана с ее исторической целостностью. Поэтому, чтобы понять данный контекст, необходимо представлять картину шире, включая доступные ресурсы и исторические и культурные аспекты.

ЗАКЛЮЧЕНИЕ

Развитие и консолидация открытого доступа происходит по-разному в развитых странах и в остальном мире, в частности в Латинской Америке. Это отличие, помимо прочего, лежит в особенностях системы научной коммуникации, сложившейся со временем в Латинской Америке. С одной стороны, эта система не обладала способностью улучшить научное развитие, а с другой (открытый доступ, особенно золотая дорога благодаря малому количеству престижных научных коммерческих издательств) – значительно перспективна в регионе. Этот вывод исходит из успешных инициатив, таких как SciELO и широко распространенная адаптация системы журналов открытого доступа. Опыт SciELO приводится в одном из документов Будапештской инициативы открытого доступа, показывая, что инициатива была передовой и повлияла на развитие открытого доступа.

Несмотря на этот латиноамериканский ракурс, консолидация золотой дороги в Латинской Америке в силу некоммерческой природы большинства научных издателей в регионе не была достаточной для процветания зеленой дороги. Сегодня мы полагаем, что поскольку большая часть научных издателей не обращает внимания на маркетинговую логику, то большинство научных журналов не ставят барьеры при повторном распространении ими научной литературы в альтернативных средствах, таких как институциональные хранилища.

Мы наблюдали, что развитие институциональных хранилищ в некоторых странах региона сопровождалось дебатами и законопроектами в общественной сфере, и таким образом параллель инициатив открытого доступа в Латинской Америке может проследиваться и в отношении других регионов мира. Однако, несмотря на недавние улучшения зеленой дороги, в частности рост числа институциональных хранилищ и усиление кооперативной работы в регионе, все еще предстоит достичь многого в этой области.

ЛИТЕРАТУРА

1. *Meadows J.* A comunicação científica. — Brasília: Briquet Lemos, 1999.
2. *Chan L., Costa S.* Participation in the global knowledge commons: Challenges and opportunities for research dissemination in developing countries// *New Library World*.— 2005. — Vol. 106, No. 3/4. — P.141-163. — <http://www.emeraldinsight.com/doi/full/10.1108/03074800510587354>
3. *Alperin J.P., Fischman G.E., Willinsky J.* Open access and scholarly publishing in Latin America: Ten flavours and a few reflections// *Liinc em Revista*. — 2008. — Vol. 4, No.2. — P.172-185. — <http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/269/167>
4. *Babini D.* Acceso abierto a la producción científica de América Latina y el Caribe: Identification of main institutions for regional integration strategies// *Revista Iberoamericana de Ciencia Tecnología y Sociedad*. —2011. — Vol.6, No.17.— P. 31-56.— <http://hdl.handle.net/10760/15574>
5. *Directory of Open Access Journals.* DOAJ Statistics. [s.l.]. — 2013. — <https://doaj.org/>
6. *Lima T., Mioto R.* Procedimentos metodológicos na construção do conhecimento científico: a pesquisa bibliográfica// *Revista Katálysis*. — 2007.— Vol.10, No.Esp.— P.37-45.—<http://www.scielo.br/pdf/rk/v10nspe/a0410spe.pdf>

7. *Mebo L.I., Yang K.* A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar// *Journal of the American Society for Information Science and Technology*. — 2006.— Vol.58, No. 2.— P.297-301.— <http://arxiv.org/abs/cs/0612132>

8. *Harnad S.* A subversive proposal/ Okerson A., O'Donnell J. (Ed.). *Scholarly journals at the crossroads: A subversive proposal for electronic publishing*. — Washington: Association of Research Libraries, 1994. — <http://www.arl.org/scomm/subversive/toc.html>

9. *Harnad S.* Research access, impact and assessment// *Times Higher Education Supplement*. — 2001. — No.18. — <http://eprints.soton.ac.uk/255950/>

10. *Odlyzko A.* The rapid evolution of scholarly communication// *Learned Publishing*.— 2002.— Vol.15, No.1. — P.7-19. —<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.115.7566>

11. *Sarmiento M. et al.* Algumas considerações sobre as principais declarações que suportam o movimento Acesso Livre/ *World Congress on Health Information and Libraries*, 9., Salvador, 2005. *Anais eletrônicos...* — Salvador: Ministério da Saúde, 2005. — <http://hdl.handle.net/10760/8512>

12. *Askey D.* Moving beyond the MLA & deviant modes of scholarly communication/ *Women in German Conference*, 2006, Snowbird. *Proceedings eletrônicos...* — Snowbird: Kansas State University, 2006. — <http://hdl.handle.net/2097/701>

13. *Civallero E.* Open access: Experiencias latinoamericanas/ *Congreso Internacional de Bibliotecología e Información*, 2., 2006, Lima. *Anales eletrônicos...*— Lima: Colegio de Bibliotecólogos del Peru, 2006. — <http://hdl.handle.net/10760/11621>

14. *Bailey C.* Open access and libraries// *Collection Management*. — 2007. — Vol. 32, No. 3-4. — P. 351-383. — <http://digitalscholarship.org/cwb/OALibraries2.pdf>

15. *Murray-Rust P.* Open data in science// *Serials Review*. — 2008. — Vol.24, No.1. — P.52-64. — <http://hdl.handle.net/10101/npre.2008.1526.1>

16. *Barros M.* A primavera acadêmica e o custo do conhecimento// *Liinc em Revista*. —2012. — Vol. 8, No. 2. — P.365-377. — <http://revista.ibict.br/liinc/index.php/liinc/article/view/486>

17. *Budapest Open Access Initiative.* Ten years on from the Budapest Open Access Initiative: Setting the default to open. — Budapeste: BOAI, 2012. — <http://www.opensocietyfoundations.org/openaccess/boai-10-recommendations>

18. *Mishra K.* Internet a boon for researchers: A study on open access resources// *Pragyaan: Journal of Mass Communication*. — 2012. — Vol.10, No.1. — P.14-22. — http://ims.edu.in/download-pdf/MJ4077_Pragyaan-MassCommunication.pdf#page=20

19. *Suber P.* What is open access? / Suber, P. (Org.). *Open access*. — Cambridge: MIT Press, 2012. — P.1-28.

20. *Budapest Open Access Initiative.* Read the Budapest open access initiative. —Budapeste, 2002. — <http://www.opensocietyfoundations.org/openaccess/read>

21. *Acordo: declaração de Santo Domingo*. — Santo Domingo, 2002. — <http://segib.org/reuniones/files/2002/06/IVConferencia-Ibero-americana-de-Ministros-da-Administracao-Publica-e-Reforma-do-Estado-Acordo.pdf>

22. *Ramlo S.* ArXiv.org and physics education// *The Physics Teacher*. —2007. — Vol. 45, No. 6. — P. 374-375. — <http://scitation.aip.org/content/aapt/journal/tpt/45/6/10.1119/1.2768698>

23. Lagoze C., Van de Sompel H. The open archives initiative: Building a low-barrier interoperability framework/ ACM/IEEE Joint Conference on Digital Libraries, 1., Roanoke, 2001. Proceedings electronics... — Roanoke: Open Archives Initiative, 2001. — P. 54-62. — <http://www.openarchives.org/documents/jcdl2001-oai.pdf>
24. Pontika N. The influence of the national institutes of health public-access policy on the publishing habits of principal investigators. Thesis (Doctorate in Philosophy). — School of Library and Information Science, Simmons College, Boston, 2011. — <http://hdl.handle.net/10760/17208>
25. Bethesda statement on open access publishing: Meeting on open access publishing. — Bethesda, 2003. — <http://legacy.earlham.edu/~peters/fos/bethesda.htm>
26. Berlin declaration open access to knowledge in the sciences and humanities. — Berlin: Open Access, 2003. — <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>
27. Valparaiso declaration for improved scientific communication in the electronic medium. — Valparaiso, 2004. — <http://eprints.rclis.org/20069/>
28. Washington DC principles for free access to science. — Washington, 2004. — <http://www.dcpinciples.org/>
29. Brussels declaration on open access. — Brussels, 2012. — <http://openaccess.be/2012/10/22/brussels-declaration-on-open-access/>
30. Croatian open access declaration. — Zagreb, 2012. — <http://www.fer.unizg.hr/oa2012/declaration>
31. Scottish universities declaration on open access. — Edinburgh, 2005. — <http://scurl.ac.uk/wp-content/uploads/2013/07/ScottishDeclarationOnOpenAccess.pdf>
32. Organisation for Economic Co-operation and Development. Principles and guidelines for access to research data from public funding. — Paris, 2007. — <http://www.oecd.org/sti/sci-tech/38500813.pdf>
33. International Federation of Library Associations and Institutions. Statement on open access: Clarifying IFLA's position and strategy. — Haia, 2011. — <http://www.ifla.org/files/assets/hq/news/documents/ifla-statement-on-openaccess.pdf>
34. Communia International Association on the Public Domain. Position on ec horizon 2020 open access policy. — Brussels, 2012. — <http://www.communia-association.org/2012/11/20/position-on-ec-horizon-2020-open-accesspolicy/>
35. Recommendation for open access and online content in LILACS. — São Paulo, 2012. — <http://metodologia.lilacs.bvsalud.org/php/level.php?lang=en&component=34&item=55>
36. Instituto Brasileiro de Informação em Ciência e Tecnologia. Manifesto brasileiro de apoio ao acesso livre à informação científica. — Brasília: IBICT, 2005. — <http://livroaberto.ibict.br/docs/Manifesto.pdf>
37. Declaração de Salvador sobre o acesso aberto: A perspectiva dos países em desenvolvimento. International Seminar on Open Access. — Salvador, 2005. — <http://www.icml.org/channel.php?lang=pt&channel=86&content=428>
38. Declaración de Cuba en favor del acceso abierto / Congreso de la Unión Latinoamericana de Entidades de Psicología, 2., 2007, Havana. Anales electrónicos... — Havana: Sociedad Cubana de Psicología de la Salud, 2007. — <http://promociondeeventos.sld.cu/psicosalud/declaracion-decuba-en-favor-del-acceso-abierto/>
39. Brasil. Projeto de lei 1.120, 21 de maio de 2007. Dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências. — 2007. — <http://www.camara.gov.br/proposicoesWeb/fichadetramitacao?idProposicao=352237>
40. Brasil. Projeto de lei do Senado 387, 5 de julho de 2011. Dispõe sobre o processo de registro e disseminação da produção técnico-científica pelas instituições de educação superior, bem como as unidades de pesquisa no Brasil e dá outras providências. — 2011. — http://www.senado.gov.br/atividade/Materia/Detalhes.asp?p_cod_mate=101006
41. Argentina. Resolución Ministerial n°469, de 17 de maio de 2011. Aprova a revisão 4 dos módulos 1 e 8 e a revisão 5 do módulo 6 dos procedimentos de distribuição de energia elétrica no sistema elétrico nacional. — http://www.biblioteca.mincyt.gob.ar/docs/res_be_469-11.pdf
42. Jiménez T. Marco de políticas de acceso abierto para la Red Mexicana de Repositorios Institucionales. — Red Mexicana de Repositorios Institucionales. Ciudad De México, 2013. — http://www.remeri.org.mx/remeri/REMEDI_POLITICAS_2013_01.pdf
43. Red Federada Latinoamericana de Repositorios Institucionales de Documentación Científica en América Latina. Inicia el cambio regional: trabajo conjunto para el acceso abierto. — 2014. — <http://lareferencia.redclara.net/rfr/noticias/inicia-el-cambio-regional-trabajo-conjunto-para-el-accesoabierto>
44. Peru. Ley 33.035. Regula el Repositorio Nacional de Ciencia, Tecnología e Innovación de Acceso Abierto. — Lima, 2013. — <http://www.unesco.org/new/en/communicationand-information/portals-and-platforms/goap/access-byregion/latin-america-and-the-caribbean/peru/>
45. Suber P. Open access policy options for funding agencies and universities. //SPARC Open Access Newsletter, n.130, 2009. — <http://www.earlham.edu/~peters/fos/newsletter/02-02-09.htm>
46. Zúñiga A. Historia de la ciencia y la tecnología y la realidad de America Latina //Elementos. —1986. — Vol. 6, No. 1. — <http://www.elementos.buap.mx/num06/pdf/17.pdf/>

Представление фасетометрии: новое измерение на основе ключевых слов для расположения научных областей предмета*

Калипада ДЖАНА
(Kalipada JANA)

Мемориальная библиотека им. Б. Ч. Роя,
Индийский институт менеджмента,
г. Калькутта, шт. Западная Бенгалия, Индия

Бидьярти ДУТТА
(Bidyarthi DUTTA)

Отделение библиотекведения и информа-
тики, Видьясагарский университет, г. Ми-
нинишур, шт. Западная Бенгалия, Индия

Вводится технология поиска научных областей в предмете с помощью анализа ключевых слов (КС), приписанных к статьям. Собирались и анализировались КС, присвоенные 1227 научным статьям, опубликованным в 2004-2013 гг. по конкретной предметной области – «Радиация Хокинга». Статьи извлекались из БД Web of Science с использованием поискового термина «Радиация Хокинга». Присвоенные КС встречались с различной частотой. Они формировали кластеры слов. Названия, даваемые кластерам, соответствовали обозначениям самых часто встречающихся слов в кластерах КС. Кластеры классифицировались по трем размерным группам: малый кластер, средний кластер и большой кластер. Поскольку большое число КС формировало большие кластеры, было высказано предположение, что потенциальные фасеты будут представлены такими кластерами. Были определены три основных параметра, ассоциируемые с кластерами КС: число КС в кластере, частота встречаемости и заполняемость. Четыре показателя, а именно, индекс стабильности, индекс интегрированной видимости, индекс моментальной видимости и индекс потенциальности, определялись на основе этих трех параметров и отмечались их распределенности в период исследования. Эти показатели сохраняют разные значения для различных кластеров и фасетов. Их ранги значимости категоризируются по пяти группам: очень высокий, высокий, средний, низкий и очень низкий. Каждый показатель описывает отдельный аспект фасета. Эти показатели могут быть использованы для измерения различных аспектов фасетов.

ВВЕДЕНИЕ

Определение научных областей является базовым требованием исследования предмета. Как указывает Ситарам [1], изучение развития предмета включает три основных компонента: **1.** вехи в развитии предмета, т.е. **2.** научные направления, взятые из обзорных документов, индексируемой и реферлируемой периодики, отчета о современной ситуации, отчета о тенденциях и т.д. Также должны быть сделаны попытки по установлению в целом (а) роста литературы (документов) по интере-

суемому предмету и (б) степени фильтрации и рассеяния информации в интересующем предмете. **3.** Направления в образовании по интересующему предмету на различных уровнях. Анализ основных направлений, как правило, превращается в фокусирование на фасетах, служащих направлениями областей интересующего предмета. В статье предпринимается попытка разработать количественный метод поиска фасетов, содержащих направления научных областей предмета «Радиация Хокинга». Этот количественный метод включает отбор КС [2] из научных статей, сопровождаемый анализом частоты. Анализ частоты отмечает кластеры КС, показывающие развивающиеся научные темы предмета. Данное этому исследованию название – *фасетометрия*, количественное исследование фасетов. Это исследование выявляет растущие и снижающиеся фасеты предметной области.

* Перевод Jana K., Dutta B. Introducing facetometrics: A new keyword-based measure for locating research areas of a subject.— <http://eprints.rclis.org/29513/1/Paper-SRELS-Jana-Dutta.pdf>

СВЯЗАННЫЕ С ИССЛЕДОВАНИЕМ РАБОТЫ

Прайс [3,4] исследовал рост числа ученых, научных журналов и статей за последние два столетия. Он обнаружил, что эти цифры удваиваются каждые 15 лет. Предполагается, что литература и информация растут экспоненциально, но в отдельных дисциплинах рост может быть линейным. Понятие старения или устаревания тесно связано с ростом науки. Некоторые авторы считали рост и устаревание взаимозаменяемыми, предполагая, что чем быстрее растет литература в области, тем быстрее она стареет и литература устаревает за более короткое время [5]. Это также может быть сделано с помощью анализа роста и старения ключевых слов в предметной области. Имеются исследования роста и старения литературы, использующие также анализ ссылок. Контент-анализ служит другим научным методом, который широко применяется в библиотекведении и информатике в целях изучения современной ситуации по предмету. Контент-анализ, впервые примененный в 1950-х гг., является системным и строгим подходом к анализу научных статей, собранных в ходе исследования [6]. Фундаментальными составляющими содержания предмета являются КС. И контент-анализ с помощью ключевых слов основан на предположении, что КС статьи дают адекватное представление о ее содержании. Два различных КС, встречающихся вместе в одной статье, служат показателем связи между темами, к которым они относятся [7]. Култер, Монарх и Конда [8] отобрали КС, выбранные профессионалами-индексаторами. Они считали, что это полезно для изучения фиксированной системы, которая налагает общую номенклатуру. Опыт профессионалов-индексаторов предполагает стандартное применение такой таксономии. Луз и Лемари [9] провели исследование совместных слов, основанное на КС, предложенных экспертами. Куртьаль, Калик и Каллон [10] загрузили КС из онлайн баз данных, чтобы исследовать социальное взаимодействие. Лоу и Уайттейкер [11] отобрали исследование окисления с помощью анализа совместно встречаемых слов. Нойон и Ван Раан [12] представили подтверждения применений в процедуре отображения науки и техники. Улучшения связаны с использованием графических интерфейсов пользователей и добавлением информации «внешнего отображения». Данный интерфейс позволяет пользователям отображений (карт) фокусироваться на их отдельных областях интереса и определять положение деятелей в области. Нойон и Ван Раан [13] отобрали целую структуру нейронных сетей, используя совместную встречаемость классификационных кодов. Ван Раан и Тейссен [14] обсуждали ограничения и возможности библиометрического отображения на основе анализа совместных слов. Ван Раан [15] анализировал основные научные области, покрываемые журналом *Scientometrics*, применяя анализ совместных слов. Куртьаль [16] вычислил взаимодействие сети внутри всевозможных авторов при помощи анализа совместных слов для журнала *Scientometrics*. Для анализа временных вариаций содержания области была разработана модель КАСОВА (Keyword associated content variation analysis – анализ вариаций содержания, связанных с КС), которая включает подробное, детальное реферирование и анализ КС [17,18]. Ключевые слова по предмету Ферми-жидкость были категоризированы как ключевая фраза, модулятор и квалификатор и затем проанализированы [18]. Для анализа научных направлений и выявления потенциальных областей исследования предмета была предложена новая таксономия КС [19].

Осознанно или неосознанно мы всегда используем «ключевое слово» в повседневности. Термин «ключевое слово в контексте» был введен Луном [20] в 1960 г. Раннее использование КС найдено в 1975 г. в журнале *Journal of Applied Behaviour Analysis* [21]. Ключевые слова представляют контекст, а анализ кластеров КС превращается в анализ контекста предмета (контекстуальный анализ предмета). По мере сужения предмета только слова, которые будут описывать предмет, становятся такими словами, с помощью которых ученые определенной области будут относить ее к данному предмету. Это подтверждает то, что Лун называл «ключевым словом» или значимым словом [22].

Анализ кластеров КС служит эффективным методом для изучения точки зрения пользователя относительно информационного пространства с целью создания гибкой и доступной классификационной схемы. Он основан на статистическом анализе различных характеристик КС. Анализ кластеров широко применяется во всех основных дисциплинах и в частности анализ кластеров на основе документов прокладывает дорогу к автоматической классификации [23].

Главным недостатком существующих методов доступа к предмету является то, что они умалчивают о поведенческих аспектах КС, т.е. формах встречаемости КС в базе данных, полнотекстовой или библиографической, где бы она не находилась. Одной из сильных сторон предлагаемой здесь модели служит количественная интерпретация поведенческих аспектов КС для изучения роста и старения различных областей предмета.

ЦЕЛИ, МЕТОДОЛОГИЯ И ОГРАНИЧЕНИЯ

Основная цель статьи заключается в развитии метода определения научных областей предмета из первоначальных источников информации, которые отражают актуальную ситуацию по предмету. Всего из БД Web of Science было извлечено 1227 научных статей по теме «Радиация Хокинга», опубликованных в период с 2004 по 2013 г. Предполагается, что поиск термина «Радиация Хокинга» отражает конкретную предметную область в рамках широкой сферы астрофизики. Затем 8592 КС были отобраны из названий и рефератов к этим 1227 статьям, т.е. в среднем семь КС на статью (табл. 1).

На рис. 1 и 2 представлены изменения (вариации) статей и КС по годам. Все эти КС затем были разбиты на две группы, а именно «Ключевые слова кластера» и «Ключевые слова без кластера». Для определения КС кластера сначала выявлялись КС с некоторыми общими словами, и они, а также группа КС со специфическими в них словами считались кластером. Кластеры категоризировались по размеру, т.е. числу КС в кластере. Кластеры, содержащие от трех до пяти КС, классифицировались как «малые кластеры»; кластеры с шестью и десятью КС – «средние кластеры», а кластеры, включающие более десяти КС, относились к «большим кластерам». Кластеризация КС означает группировку одного и того же в общий фасет. Здесь термин «фасет» отражает кластер КС с общим словом. Каждый кластер именуется по общему слову. Некоторые примеры кластеров КС приведены в табл. 2.

Общее число отобранных КС за десять лет (табл. 1) составляет 3509, включая повторы одного и того же КС в разные годы. Любое узкое КС может встречаться в разных годах. Число различных ключевых слов за десять лет, т.е. без повторов, равнялось 1630. Значит, в среднем каждое КС повторялось ($8592/1630= 5,3$) примерно пять раз в течение десяти лет. В среднем три КС отбирались из каждой статьи за десять лет ($3509/1227= 2,9$), также имеет-

ся 1,3 (1630/1227=1,3) различных КС на статью. Из всех 1630 различных КС 796 КС встречались без повторов. Эти 796 КС не образовали никакого кластера и именовались «ключевыми словами без кластера». Средняя частота встречаемости КС без кластера равняется 3,7 (2965/796=3,7). Оставшиеся 834 ключевых слова встречались 5627 раз и сформировали 84 кластера. Из этих 84 кластеров 48 являются малыми, 25 кластеров – средними и 11 кластеров – большими по размеру. Поскольку большие кластеры содержат большее число релевантных КС, то они могут рассматриваться как «ядерные» дескрипторы или основные идеи предмета. Точно так же кластеры среднего и малого размера могут считаться «близкими» дескрипторами и «чуждыми» дескрипторами соответственно. Интересно, что число статей демонстрирует рост до 2008 г. (170), а затем начинает снижаться. Модель вариации отобранных КС была почти идентична модели числа статей.

Цифровые значения четырех характерных показателей (characteristic indicators - CI) для каждого крупного и среднего кластера были подсчитаны и представлены в табл. 3. Этими CI являются: показатель интегрированной видимости, показатель моментальной видимости, показатель стабильности и показатель потенциальности, которые определяются следующим образом [22]:

показатель интегрированной видимости, т.е. $v = F/N$,
показатель моментальной видимости, т.е. $m = F/A$,
показатель потенциальности, т.е. $p = \ln(N \cdot F)$ и
показатель стабильности, т.е. $s = (A/A_{\max}) \cdot 100$, где

N - число КС в кластере,

F - частота встречаемости всех КС кластера,
 A - встречаемость всех КС в кластере, т.е. число появлений КС за оговоренный временной период,

A_{\max} – максимальная встречаемость всех КС в определенном кластере, т.е. число КС кластера умножится на число лет в оговоренном временном периоде.

Таблица 1

Выборка для исследования

Год	Число статей (А)	Число отобранных КС (В)	Общее число встречаемостей всех КС (С)	Средняя частота (округл.) встречаемости на КС (С/В)	Число отобранных (округл.) КС на статью (В/А)	Число встречаемости КС (округл.) на статью (С/А)
2004	57	200	374	2	4	7
2005	62	195	356	2	3	6
2006	109	338	706	2	3	6
2007	145	370	934	3	3	6
2008	170	406	1145	3	2	7
2009	165	433	1252	3	3	8
2010	151	431	1182	3	3	8
2011	127	403	912	2	3	7
2012	124	381	911	2	3	7
2013	117	352	820	2	3	7
Всего	1227	3509	8592	2	3	7

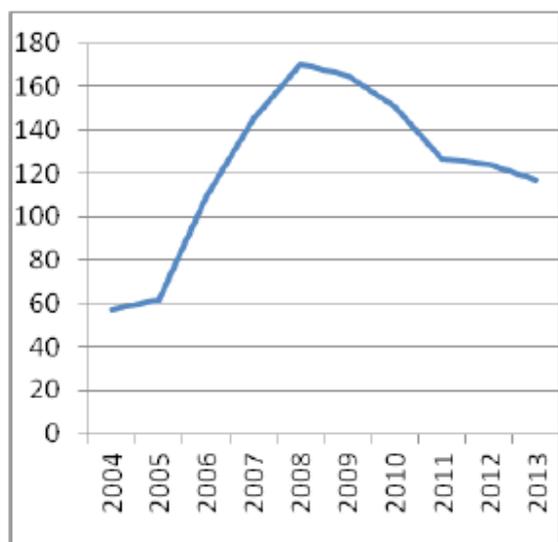


Рис. 1. Число статей за десять лет

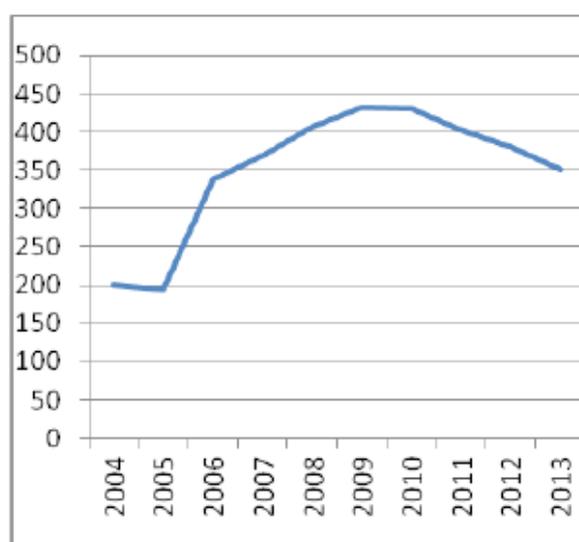


Рис. 2. Число отобранных КС за десять лет

Некоторые примеры малых, средних и больших кластеров КС

Размер кластера	Имя кластера и число КС	КС кластера
Малый	Аккреция (4)	Аккреция, аккреция диска, эффективность аккреции, адвективно-доминирующая аккреция
Малый	Темная энергия (4)	Темная энергия, теория темной энергии, голографическая темная энергия, графическое изображение темной энергии по возрасту
Средний	Брана (6)	Брана, Дирихле-брана (D-брана), подход D-бран, производство бран, мир на бранах, сценарий мира на бранах
Средний	Кварк (6)	Странная кварковая материя, физика тяжелых кварков, производство кварков, кварковая звезда, кварк глюонной плазмы, кварковая струя
Средний	Термодинамика (9)	Геометрия термодинамики, свойство термодинамики, количество термодинамики, термодинамика, общий закон термодинамики, первый закон термодинамики, термодинамика вселенной Фридмана-Робертсона-Уолкера, второй общий закон термодинамики, термодинамика общего принципа неопределенности
Большой	Энтропия (27)	Сила энтропии, акустическая неустойчивость энтропии, энтропия и область квантования, скачок энтропии, сохранение энтропии, коррекция энтропии, коррекция энтропии Бекенштейна-Хокинга, плотность энтропии, энтропия горизонта, квантование энтропии, энтропия спектра, энтропия Бекенштейна-Хокинга, энтропия Хокинга, каноническая энтропия, скорректированная энтропия, коррекционная энтропия, запутанность энтропии, энтропия Ферми, геометрическая энтропия, гравитационная энтропия, энтропия дыр, энтропия логарифмов, энтропия коррекции логарифмов, модифицированная энтропия, абсолютная энтропия Планка, полуклассическая энтропия

Таблица 3

Характерные показатели для ядерных дескрипторов

Крупные кластеры КС (ядерные фасеты)	Число КС в кластере, N)	Частота встречаемости, F)	(Наполняемость, A)	[Максимальная наполняемость, A(max)]	Показатель интегрированной видимости, $v = F/N$	Показатель моментальной видимости, $m = F/A$	Показатель потенциальности, $p = \ln(N*F)$	Показатель стабильности, $S = (A/A(max))*100$
Черная дыра	154	832	301	1540	5,40	2,76	11,76	19,55
Космология	24	85	50	240	3,54	1,70	7,62	20,83
Теория Эйнштейна	11	19	17	110	1,73	1,12	5,34	15,45
Энтропия	27	321	64	270	11,89	5,02	9,07	23,70
Гравитация	51	300	130	510	5,88	2,31	9,64	25,49
Квантовая физика	61	276	136	610	4,52	2,03	9,73	22,30
Радиация	11	110	32	110	10,00	3,44	7,10	29,09
Область масштабирования	11	36	27	110	3,27	1,33	5,98	24,55
Пространство-время	36	150	75	360	4,17	2,00	8,59	20,83
Туннелирование	22	158	63	220	7,18	2,51	8,15	28,64
Вселенная	13	75	46	130	5,77	1,63	6,88	35,38

Следует отметить, что частота встречаемости отличается от числа появлений КС. Предположим, КС находится в базе данных в течение трех лет, т.е. 2011 г., 2012 г. и 2014 г. с частотой 12, 14 и 15. Значит, общая частота встречаемости этого КС за три года складывается $12+14+15=41$, а число его появлений равняется трем, т.е. число появлений или наполняемость просто подсчитывается по тому, как много лет оно встречалось. Максимальная наполняемость отражает самое высокое

появление КС в течение временного периода. Например, с 2001 по 2015 гг., если КС встречалось только в 2005, 2007 и 2010 гг., то его наполняемость (A) должна быть равной трем, тогда как максимальная наполняемость (A_{max}) должна быть 15, т.е. протяженность всего временного периода.

ПРЕДСТАВЛЕНИЕ В ВИДЕ СХЕМЫ

Шаги, включенные в функцию *фасетометрии*, схематично представлены на рис. 3.

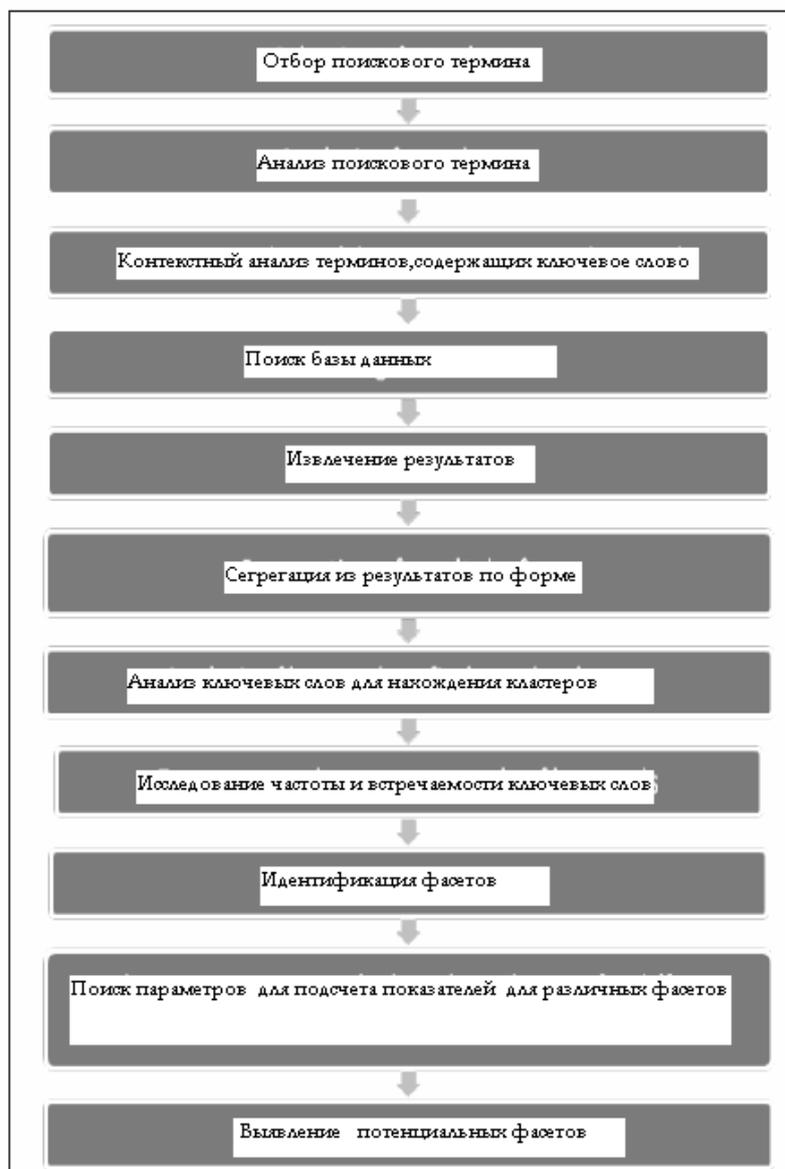


Рис. 3. Шаги, включенные в фасетометрию

РЕЗУЛЬТАТЫ И АНАЛИЗ

В ходе этого исследования найдены четыре характерных показателя для ядерного и чужого фасетов. Показатель интегрированной видимости (v) фасета или кластера определяется как средняя частота встречаемости на КС. Показатель моментальной видимости (m) фасета или кластера определяется как средняя частота встречаемости на КС для единицы появления или наполняемость, т.е. как часто в среднем КС приписывается в научных статьях в течение года. Показатель потенциальности (p) определяется как натуральный логарифм произведения общего числа КС и частоты. Таким образом он служит показателем и числа КС, и их частот фасета, которые могут считаться силой фасета. Показатель стабильности определяется как соотношение реальной наполняемости к максимальной наполняемости фасета. Он отражает временную стабильность или стабильность фасета за временной период и показывает, появляются ли КС фасета равномерно за целый временной период. Всего наблюдалось 11 ядерных дескрипторов, 25 близких дескрипторов и 48 чужих дескрипторов. Соотношение числа дескрипторов в трех зонах состав-

ляет 11: 25: 48 или 11 (1: 2,3: 4,4). Это соотношение почти совпадает с моделью Брэдфорда, т.е. $k(1:n:(n^2))$, где k – множитель Брэдфорда. Здесь $k=11$, $n=2,3$ и $(n)^2=4,4$ (близко к реальному значению, т.е. 5,3). Значения характерных показателей для ядерных и близких дескрипторов показано в табл. 3 и 5 соответственно. Все дескрипторы ранжированы в соответствии с цифровыми значениями каждого из четырех показателей. Ранжирование ядерных дескрипторов и близких дескрипторов представлено в табл. 4 и 6 соответственно. Для сравнения четырех ранжирований использовался коэффициент корреляции ранжирований Спирмана, и эти данные отражены в табл. 7. Отмечается, что все коэффициенты для близких дескрипторов обладают сильной положительной корреляцией, и только три коэффициента для ядерных дескрипторов имеют сильную положительную корреляцию, два коэффициента – слабую положительную корреляцию и один – слабую отрицательную корреляцию. Ранжирование ядерных дескрипторов (табл. 4) говорит о том, что следующие пять дескрипторов, а именно Черная дыра, Энтропия, Квантовая физика, Радиация и Вселенная, составляют

центральные фасеты предмета. Однако ранжирования в соответствии с показателем стабильности и показателем потенциальности в случае ядерных дескрипторов не находятся в гармонии, а почти наоборот. Это означает, что в ядерных областях потенциальные фасеты выглядят менее стабильными во времени и что временно стабильные дескрипторы содержат сравни-

тельно меньшее число КС. Взгляд на табл. 4 проясняет эту особенность, например, «Черная дыра» и «Квантовая физика» ранжированы под «1» и «2» показателем потенциальности, тогда как показателем стабильности эти два дескриптора ранжированы как «7» и «10». Ранжирование близких дескрипторов всеми показателями выдает приблизительно идентичные результаты.

Таблица 4

Ранжирование ядерных дескрипторов

Ранг ядерных фасетов СИ	Ранжирование в соответствии с значениями четырех СИ			
	Показатель интегрированной видимости (v)	Показатель моментальной видимости (m)	Показатель потенциальности (p)	Показатель стабильности (s)
1	Энтропия	Энтропия	Черная дыра	Вселенная
2	Радиация	Радиация	Квантовая физика	Радиация
3	Туннелирование	Черная дыра	Гравитация	Туннелирование
4	Гравитация	Туннелирование	Энтропия	Гравитация
5	Вселенная	Гравитация	Пространство-время	Область масштабирования
6	Черная дыра	Квантовая физика	Туннелирование	Энтропия
7	Квантовая физика	Пространство-время	Космология	Квантовая физика
8	Пространство-время	Космология	Радиация	Пространство-время
9	Космология	Вселенная	Вселенная	Космология
10	Область масштабирования	Область масштабирования	Область масштабирования	Черная дыра
11	Теория Эйнштейна	Теория Эйнштейна	Теория Эйнштейна	Теория Эйнштейна

Таблица 5

Характерные показатели (СИ) для близких дескрипторов

Средние кластеры КС (близкие фасеты)	Число КС в кластере, N	Частота встречаемости, F	Наполняемость, A	Максимальная наполняемость, A(max)	Показатель интегрированной видимости, v=F/N	Показатель моментальной видимости, m=F/A	Показатель потенциальности, p=ln(F*N)	Показатель стабильности, S=(A/A(max))*100
Аномалия	7	138	27	70	19,71	5,11	6,87	38,57
Область спектра	6	44	15	60	7,33	2,93	5,58	25,00
Теория бран	6	70	24	60	11,67	2,92	6,04	40,00
Теория сохранений	6	7	7	60	1,17	1,00	3,74	11,67
Координатная трансформация	8	13	12	80	1,63	1,08	4,64	15,00
Теория Дирака	7	124	26	70	17,71	4,77	6,77	37,14
Электромагнитная область	8	10	11	80	1,25	0,91	4,38	13,75
Энергия	10	84	30	100	8,40	2,80	6,73	30,00
Теория измерений	9	78	22	90	8,67	3,55	6,55	24,44
Горизонт	9	146	18	90	16,22	8,11	7,18	20,00
Информация	9	59	29	90	6,56	2,03	6,27	32,22
Физика частиц	9	210	39	90	23,33	5,38	7,54	43,33
Пертурбация	7	39	9	70	5,57	4,33	5,61	12,86
Распределение Планка	6	21	13	60	3,50	1,62	4,84	21,67
Кварк	6	7	6	60	1,17	1,17	3,74	10,00
Квази-функция	6	49	16	60	8,17	3,06	5,68	26,67
Относительность	6	28	16	60	4,67	1,75	5,12	26,67
Полуклассический подход	6	9	8	60	1,50	1,13	3,99	13,33
Пространство	8	95	26	80	11,88	3,65	6,63	32,50
Статистический подход	6	24	15	60	4,00	1,60	4,97	25,00
Теория струн	6	77	26	60	1,28	2,96	8,44	43,33
Термодинамика	9	177	20	90	19,67	8,85	7,37	22,22
Принцип неопределенности	9	59	20	90	6,56	2,95	6,27	22,22
Волна	7	17	13	70	2,43	1,31	4,78	18,57
Приближение ВКБ	6	17	14	60	2,83	1,21	4,62	23,33

Ранжирование чужих дескрипторов

Ранг чужих фасетов СИ	Ранжирование чужих фасетов в соответствии с значениями четырех СИ			
	Показатель интегрированной видимости (v)	Показатель моментальной видимости (m)	Показатель потенциальности (p)	Показатель стабильности (s)
1	Физика частиц	Термодинамика	Теория струн	Теория струн
2	Аномалия	Горизонт	Физика частиц	Физика частиц
3	Термодинамика	Физика частиц	Термодинамика	Теория бран
4	Теория Дирака	Аномалия	Горизонт	Аномалия
5	Горизонт	Теория Дирака	Аномалия	Теория Дирака
6	Пространство	Пертурбация	Теория Дирака	Пространство
7	Теория бран	Пространство	Энергия	Информация
8	Теория измерений	Теория измерений	Пространство	Энергия
9	Энергия	Квази-функция	Теория измерений	Квази-функция
10	Квази-функция	Теория струн	Принцип неопределенности	Относительность
11	Область спектра	Принцип неопределенности	Информация	Область спектра
12	Информация	Область спектра	Теория бран	Статистический подход
13	Принцип неопределенности	Теория бран	Квази-функция	Теория измерений
14	Пертурбация	Энергия	Пертурбация	Приближение ВКБ
15	Относительность	Информация	Область спектра	Термодинамика
16	Статистический подход	Относительность	Относительность	Принцип неопределенности
17	Распределение Планка	Распределение Планка	Статистический подход	Распределение Планка
18	Приближение ВКБ	Статистический подход	Распределение Планка	Горизонт
19	Волна	Волна	Волна	Волна
20	Координатная трансформация	Приближение ВКБ	Координатная трансформация	Координатная трансформация
21	Полуклассический подход	Кварк	Приближение ВКБ	Электромагнитная область
22	Теория струн	Полуклассический подход	Электромагнитная область	Полуклассический подход
23	Электромагнитная область	Координатная трансформация	Полуклассический подход	Пертурбация
24	Теория сохранения	Теория сохранения	Кварк	Теория сохранения
25	Кварк	Электромагнитная область	Теория сохранения	Кварк

Таблица 7

Коэффициенты ранга корреляции между различными СИ для близких и чужих дескрипторов

Коэффициенты ранга корреляции (r) между СИ	Значения r для ядерных кластеров	Характер корреляции	Значения r для близких кластеров	Характер корреляции
r_{vm}	0,864	Strong + ve	0,867	Strong + ve
r_{vp}	0,391	Weak + ve	0,791	Strong + ve
r_{vs}	0,627	Strong + ve	0,623	Strong + ve
r_{mp}	0,645	Strong + ve	0,892	Strong + ve
r_{ms}	0,236	Weak + ve	0,529	Strong + ve
r_{ps}	-0,173	Weak + ve	0,729	Strong + ve

Характер корреляции $1 \geq r \geq 0,5$ показывает корреляцию Strong+ve,
 $0,5 > r > 0$ показывает корреляцию Weak + ve,
 $r = 0$ указывает на отсутствие корреляции,
 $0 > r \geq -0,5$ показывает корреляцию Weak – ve,
 $-0,5 > r \geq -1$ показывает корреляцию Strong – ve.

ЗАКЛЮЧЕНИЕ

В этом исследовании предложен метод количественной оценки характеристик фасетов предмета. Так как поведенческие аспекты фасетов измеряются в условиях цифровых показателей, то этот метод назван «фасетометрией». Характерные показатели подразумевают раз-

личные свойства фасетов, которые в свою очередь обуславливают поведение предмета. Например, для «Радиации Хокинга» найдено, что фасеты, сформированные временно стабильными дескрипторами, состоят из меньшего числа КС с низкой частотой, что может быть особым поведением этого предмета. Однако это становится ясным после применения данного метода к другим предметам. Данное исследование может быть распространено на другие области физических наук, а также на более широкие предметы, как бионауки, медицина, сельское хозяйство, социология, инженерия, когнитивная наука, поведенческая наука и т.д.

ЛИТЕРАТУРА

1. *Seetharama S.* Information consolidation and repackaging: Framework, methodology, planning. — New Delhi: Ess Ess Pub, 1997.
2. *Index Term.* — n.d. — https://en.wikipedia.org/wiki/Index_term
3. *de Solla Price D.J.* Little science, big science. — New York: Columbia University Press, 1963.
4. *de Solla Price D.J.* Science since Babylon. Enlarged ed. — New Haven: Yale University Press, 1975. — p. 165–186.
5. *N. d.* — Retrieved on 5 Mar 2015. — http://shodhganga.inflibnet.ac.in/bitstream/10603/5109/10/10_chapter%201.pdf
6. *Berelson B.* Content analysis in communications research. — New York: Free Press, 1952.
7. *Cambrosio A., et al.* Historical scientometrics? Mapping over 70 years of biological safety research with co-word analysis// *Scientometrics.* — 1993. — Vol. 27, No.2. — P. 119–43.
8. *Coulter N., Monarch I., Konda S.* Software engineering as seen through its research literature: A study in co-word analysis// *Journal of the American Society for Information Science.* — 1998. — Vol. 49, No.13. — P. 1206–1223.
9. *Looze M.D., Lemarie J.* Corpus relevance through co-word analysis: An application to plant proteins// *Scientometrics.* — 1997. — Vol. 39, No. 3. — P. 267–280.
10. *Courtial J.P., Cablik T., Callon M.* (A) Model for social interaction between cognition and action through a key-word simulation of knowledge growth// *Scientometrics.* — 1994. — Vol. 31, No. 2. — P. 173–192.

11. *Law J., Whittaker J.* Mapping acidification research: A test of the co-word method// *Scientometrics.* — 1992. — Vol. 23, No.3. — P. 417–461.
12. *Noyons E.C.M., van Raan A.F.J.* Advanced mapping of science and technology// *Scientometrics.* — 1998. — Vol. 41, No.1–2. — P. 61–67.
13. *Noyons E.C.M., van Raan A.F.J.* Monitoring scientific developments from a dynamic perspective: Self-organized structuring to map neural network research// *Journal of the American Society for Information Science.* — 1998. — Vol.49, No. 1. — P. 68–81.
14. *Van Raan A.F.J., Tijssen R.J.W.* The neural net of neural network research: An exercise in bibliometric mapping// *Scientometrics.* — 1993. — Vol. 26, No. 1. — P. 169–192.
15. *Van Raan A.F.J.* *Scientometrics: State-of-the-art// Scientometrics.* — 1997. — Vol. 38, No. 1. — P. 205–218.
16. *Courtial J.P.* (A) Co-word analysis of scientometrics// *Scientometrics.* — 1994. — Vol. 31, No. 3. — P. 251–260.
17. *Dutta B.* An analytical investigation into the characteristics of a subject - A case study. — PhD thesis. — Jadavpur University, 2008.
18. *Dutta B., Majumder K.P., Sen B.K.* Content management by keywords: An analytical study// *SRELS Journal of Information Management.* — 2010. — Vol. 47, No.6. — P. 599–620.
19. *Dutta B., Majumder K.P., Sen B.K.* Classification of keywords extracted from research articles published in science journals// *Annals of Library and Information Studies.* — 2008. — Vol. 55, No.4. — P. 317–333.
20. *Luhn H.P.* Keyword-in-context for technical literature (KWIC Index)// *American Documentation.* — 1960. — Vol.11.
21. *Hartley J., Kostoff R.N.* How useful are 'keywords' in scientific journals? // *Journal of Information Science.* — 2003. — Vol. 29, No. 5. — P. 433–438.
22. *Black J.D.* The keyword: It's use in abstracting, indexing and retrieving information// *Aslib Proceedings.* — 1962. — Vol.14, No.10. — P. 313–321.
23. *Willett P.* Recent trends in hierarchical document clustering: A critical review// *Information Processing and Management.* — 1988. — Vol. 24, No. 5. — P. 577–597.

Лучший из двух миров? Одновременная оценка ученых и их работ*

Эфранс Абу ЮДЖУМ
(Ephrance Abu UJUM)

Куру РАТНАВЕЛУ
(Kuru RATNAVELU)

Институт математических наук, Университет Малайи, г. Куала-Лумпур, Малайзия

Ганган ПРАТАП
(Gangan PRATHAP)

Национальный институт междисциплинарных наук и технологий, Совет научных и промышленных исследований, г. Тируванантапурам, шт. Керала, Индия

В статье исследуется двойная система оценки, которая одновременно оценивает относительную важность ученых и их работ. Она является модификацией алгоритма CITECH, недавно описанного Палом и Руджем [1]. Используя доступные библиографические записи $m \times n$ авторских ключевых слов (в качестве условного обозначения ученых) и n статей, становится возможным построить $m \times n$ матрицу свойств автор-статья. Далее она комбинируется с данными цитирований в строительстве алгоритма HITS-like, итеративно удовлетворяющего двум критериям: первый – хороший автор цитируется хорошими авторами, а второй – хорошая статья цитируется хорошими авторами. Следуя Палу и Руджу, готовый алгоритм производит собственную оценку автора (author eigenscore) и собственную оценку статьи. Данный алгоритм проверяется на 213 530 цитируемых публикациях, приведенных в списке под категорией «Информатика и библиотековедение» Указателя цитированных журналов по науке и технике Института научной информации Томсона за 1980-2012 гг.

ВВЕДЕНИЕ

Ранжирования предоставляют эффективные средства искусственного присвоения порядка постоянно растущему объему публикуемых исследований и ученых. Изучение и развитие такой работы постоянно склоняется в сторону так называемой *библиометрической аналитики*, которую мы здесь определяем как** «ключевые показатели, получаемые из библиометрических данных посредством математического или статистического анализа в целях обобщения понимания». Помимо информационного поиска библиометрическая аналитика фокусируется на открытии моделей, ориентированных на находящиеся под рукой данные, с целью поддержки принятия решения или задач, связанных с умозаключением. Статья является еще одним шагом в этом направлении.

Конкретнее, статья строится на недавно проделанной Палом и Руджем [1] работе относительно одновременной оценки научных авторов и статей по относительной важности. Предлагаемый алгоритм, дублирующий

СИТЕХ (CITation index), имеет преимущество множественного соответствия между определенным множеством авторов и статей, опубликованных ими совместно. С этой целью отображения обоих множеств могут быть формализованы как связи двудольного графа, который здесь относится к сети автор-статья (или автор-документ). В этом смысле кумулятивное преимущество, получаемое авторами благодаря их статьям и наоборот, может быть определено количественно с использованием методов теории графов.

Кроме того, статьи взаимосвязаны через ссылки цитирований, т.е. типичная статья ссылается на предыдущие работы в целях признания *релевантности** помимо обособления ее собственного расположения** внутри существующей литературы. Итоговая сеть цитирований

* Перевод Ujum E.A., Prathap G., Ratnavelu K. Best of both worlds? Simultaneous evaluation of researchers and their works.— <https://arxiv.org/pdf/1506.05404.pdf>

** Отмечаем, что Батт и Мартенс [2], а также Ретлефсен и Олдрич [3] использовали термин «библиометрическая аналитика», но не представили формального определения.

* Вообще связи цитирований создаются, чтобы скорее показать *реакцию* на раннюю работу, а не конкретную *зависимость*. Поэтому присутствие связей цитирований - т.е., связь, проводимая от цитирующей (ссылающейся) статьи к цитируемой (на которую ссылаются) статье, – служит описанию интеллектуальных потоков в успешных работах, что само по себе обязательно подразумевает поток влияния.

** Это понятие *расположения* статьи в отличие от ее *положения*. В отношении подсчета ссылок наличие связи цитирований явно не передает того, занимает ли она положение в поддержку или против работы, на которую ссылаются.

статьи таким образом может быть представлена направленным графом. Поскольку распределение ссылок цитирований изменяется от одной статьи к другой, (как правило, весьма искаженно [4-7]), это можно потом использовать в качестве основы для различия, какие статьи более заметно расположены, чем другие. Предлагались некоторые схемы для более подробного изучения этого признака; т.е. оценки вычисляются для каждой статьи на основе некоторой пристрастной функции ее связности (или того, как она встроена в структуру связей) [8]. Эти статьи затем могут быть классифицированы в соответствии с вычисленными оценками для получения ранжирования. Такие схемы интегральны по отношению к задачам информационного поиска в онлайн базах данных, например, Google Scholar, CiteSeerX и Microsoft Academic Search.

CITEX расширяет эту традицию путем сочетания информации из сети автор-статья с сетью цитирований статьи, чтобы определить, какие авторы и какие статьи выделяются больше других. Разработка таких алгоритмов важна для изучения альтернативных средств объединения библиометрических показателей (и их производных ранжирований) через целевую интеграцию доступной информации. CITEX представляет интерес своим построением, потому что он обеспечивает спаренную дуальную систему оценки: оценку относительной важности для авторов и другую – для статей; следовательно, сравнительные положения создателей знания и результаты их труда могут быть определены в одном подходе. Говоря проще, CITEX утверждает, что (1) хорошие авторы либо высоко производительны или высоко цитируются хорошими авторами; и (2) хорошие статьи делятся между одними и теми же авторами или цитируются хорошими статьями.

Статья организована следующим образом. Мы приводим детальное обсуждение строительства алгоритма CITEX в разделе «Алгоритм CITEX», а критика дается в разделе «Ожидаемое поведение и неясности». Предлагаемая нами модификация, именуемая здесь как алгоритм CAPS (сдвоенная оценка автор-статья), затем подробно описывается в следующем разделе «Улучшенный сдвоенный алгоритм оценки автор-статья». Для обеспечения пунктов сравнения оба алгоритма проверяются на реальном мировом массиве данных в разделе «Эмпирическая проверка». Он включает 200 тыс. статей + статьи, цитируемые Институтом научной информации и опубликованные в 1980-2012 гг., содержащиеся в списке Указателя цитированных журналов по науке и технике в предметной категории «Информатика и библиотекосведение». Статья завершается разделом «Заключение».

АЛГОРИТМ CITEX

Допустим, что перед нами массив, содержащий m авторов и n статей. Далее, предположим, что из этого массива мы можем извлечь бинарную матрицу свойств автор-статья $m \times n$, M , и бинарную матрицу ссылок $n \times n$, C . Пусть запись $M_{ij} = 1$ означает, что автор i в i -й строке M имеет (со)авторство статьи j в j -м столбце M (иначе, $M_{ij} = 0$). Это означает, что строка сумм M соответствует всем статьям, опубликованным каждым автором. Столбец сумм M соответствует всем авторам для каждой статьи. Нормализованная по столбцам версия M (с одинаковыми измерениями) может быть создана таким образом, чтобы доля авторства автора i в статье j делилась поровну $W_{ij} = M_{ij} / \sum_i M_{ij}$.

Аналогично, пусть $C_{ij} = 1$ означает, что цитируемая статья j в j -м столбце C получает ссылку из цитирующей статьи i в i -й строке C (иначе, $C_{ij} = 0$). Более того, мы тре-

буем, чтобы C не содержало самоцитирований ($C_{ii} = 0$). Учитывая экстремальный случай, когда $C = 0_{n \times n}$, Пал и Рудж определяют оценки статьи CITEX и автора как $y_j = \sum_{i=1}^m M_{ij} x_i$ и $x_i = \sum_{j=1}^n W_{ij} y_j$ соответственно. Эти выражения записаны в матричную форму как $y \leftarrow M^T x$ и $x \leftarrow W y$. Это содержит понимание того, что оценка y статьи j зависит от относительной важности ее авторов, тогда как оценка x автора i зависит от ее доли авторства (W_{ij}) каждой статьи j , умноженной на соответствующую ей оценку y_j .

Полное описание, тем не менее, требует включения свойств ссылок. Так как оно должно сводиться к случаю нулевой матрицы ссылок, Пал и Рудж получают его за счет включения термина $(I + C^T)$, эквивалентного добавлению в C самоцитирований в статьях. Поскольку $y \leftarrow M^T W y$ и $x \leftarrow W M^T x$, тогда для k -й рекурсии:

$$x^{(k)} = W(I + C^T)M^T x^{(k-1)} \quad (1)$$

$$y^{(k)} = (I + C^T)M^T W y^{(k-1)} \quad (2)$$

является одним таким возможным выбором. Путем индукции мы получаем:

$$x^{(k)} = [W(I + C^T)M^T]^k x^{(0)} \quad (3)$$

$$y^{(k)} = [(I + C^T)M^T W]^k y^{(0)} \quad (4)$$

Для векторов исходного приближения Пал и Рудж используют $x^{(0)} = 1_{m \times 1}$ и $y^{(0)} = 1_{n \times 1}$.

Предположим, $P = W(I + C^T)M^T$ при условии, что $x^{(k)} = P^k x^{(0)}$, тогда:

$$x^{(k+1)} = P P^k x^{(0)} = P x^{(k)} \quad (5)$$

Если расстояние между двумя векторами оценок x составляет $\|x^{(k+1)} - x^{(k)}\| < \epsilon$, тогда конвергенция удовлетворяется относительно толерантности ϵ [9]. Так как P не является отрицательной матрицей с размерами $n \times n$ и $x^{(0)} > 0$, тогда согласно теореме Перрона-Фробениуса*, оценки x становятся постоянными при $k \rightarrow \infty$, таким образом удовлетворяя $P x^* = x^*$ [10,11]. Подобный аргумент применим к y при условии $Q = (I + C^T)M^T W$.

Именуются и другие алгоритмы, сочетающие свойства автора и статьи. Одним примечательным примером служит подход совместного ранжирования, предложенный в [12]. Данный подход использует модель PageRank применительно к состоящей из двух частей сети цитирований – соавторство/статья, в которой два внутриклассовых случайных перемещения позволяют непосредственно пересекать линию между одним классом узлов, тогда как межклассовые случайные перемещения разрешают скачки между сетями. Стационарные вероят-

* В частности, учитывая, что $P x = c x$, а $c=1$ является самой большой собственной оценкой, тогда $P^{(k)} x^{(0)}$ сходится к вектору x^* (в одинаковом направлении, что и x) при $k \rightarrow \infty$.

ности авторских узлов и узлов статей вычисляются за счет связывания в пары случайных перемещений (полагая, что статус ученых и создаваемых ими работ взаимно усиливается). Итоговый алгоритм дает улучшения при сравнении с применением PageRank на любом свойстве (сети) отдельно, хотя и за счет введения трех дополнительных регулируемых параметров к обычному на основе одного параметра PageRank*. CИТЕХ добавляет интересный поворот в сторону текущей литературы, так как в отличие от PageRank он не зависит от любого регулируемого параметра.

ОЖИДАЕМОЕ ПОВЕДЕНИЕ И НЕЯСНОСТИ

Поскольку эффективность алгоритма извлечения данных зависит от его цели [15, 16], то полезно точно определить, какие свойства выделяются CИТЕХ, чтобы предугадать качественные аспекты ранжирования, которые оно обязательно создаст. В частности, нам интересны условия, максимизирующие определенную оценку, так как наиболее высокий процентиль разрабатывается для соответствия самым верхним рангам. Применительно к оценке автора CИТЕХ, уравнение 1 может быть расширено до:

$$x^k = WM^T x^{(k-1)} + WC^T M^T x^{(k-1)} \quad (6)$$

$$x_i^{(k)} = \sum_{a=1}^m \sum_{p=1}^n W_{ip} M_{ap} x_a^{(k-1)} + \sum_{a=1}^m \sum_{p_1, p_2=1}^n W_{ip_1} C_{p_2 p_1} M_{ap_2} x_a^{(k-1)} \quad (7)$$

Первый свободный член правой части уравнения (7) охватывает *кумулятивную долю авторства* автора i с автором a . Этот член позитивно смещается в направлении автора i , если он производителен (регулируем для доли авторства) и даже, если он часто сотрудничает с «хорошими авторами» (y которых высокие оценки x). Это включает случай, когда $a=i$, с тем, что если кумулятивная доля авторства самого автора i значительно больше, тогда $x_i^{(k)} \sim x_i^{(k-1)} \sum_{p=1}^n W_{ip}$.

Что касается второго члена уравнения, ссылка из статьи $p_2 \rightarrow p_1$ соответствует авторской ссылке из $a \rightarrow i$, разделенной на W_{ip_1} . Следовательно, этот член растет тем больше, чем имеется большее число ссылок из $a \rightarrow i$, чем больше доля авторства для каждой статьи, авторизированной под i (для которой доверием служит минимальное расхождение) и чем больше имеется оценок x автора i цитирующими авторами. Суммируя, CИТЕХ *определяет хорошего автора как того, кто часто публикуется с хорошими авторами, и, даже больше, если он цитируется хорошими авторами.*

Подобный анализ можно проделать и для оценки статьи CИТЕХ, как дается в уравнении (2):

$$y^{(k)} = M^T W y^{(k-1)} + C^T M^T W y^{(k-1)} \quad (8)$$

* Мы касаемся *параметра затухания*, изначально описанного Брином и Лейджем [13]. Заинтересованный читатель может обратиться к Лангвиллю и Майеру [14], а также к Чену с соавторами [8] для более детального обсуждения алгоритма PageRank.

$$y_j^{(k)} = \sum_{a=1}^m \sum_{p=1}^n M_{aj} W_{ap} y_p^{(k-1)} + \sum_{a_1, a_2=1}^m \sum_{p=1}^n C_{pj} M_{a_1 j} W_{a_2 p} y_p^{(k-1)} \quad (9)$$

Справа в уравнении (9) мы опять видим, что CИТЕХ определяет относительную важность в условиях обоих компонентов; первое условие охватывает публикационные свойства, тогда как второе – свойства цитирования.

Что касается первого члена, то видно, что статья j получает дробные вклады оценки y для каждого автора a , появляющегося в обеих статьях j и p . Существенно, что $M_{aj} W_{ap}$ является условием *сходства авторов*, а значит эта часть уравнения растет для статей, имеющих одинаковых авторов. Этот член будет также расти тем больше, чем большей будет оценка y для каждого «схожего автора» статьи p (относительно j) и всякий раз, когда $W_{ap} \rightarrow 1$. В случае автора i со значительно большим числом статей, мы могли бы выйти на $y_j^{(k)} \sim y_j^{(k-1)} \sum_{p=1}^n W_{ip}$.

Что касается второго члена, то видно, что y_j зависит от суммы оценок y из каждой статьи, которая ее цитирует, p («хорошие статьи» имеют высокую оценку y). После некоторых перестановок второй член также содержит произведение $W_{a_2 p} C_{pj} M_{a_1 j}$. Это значит, что оценка y статьи j зависит от суммы дробных ссылок из всех цитирующих статей p (т.е. $\sum_{p=1}^n W_{a_2 p} C_{pj}$). Сочетая это с влиянием первого члена уравнения (9), мы предполагаем, что CИТЕХ *определяет хорошую статью как ту, которая имеет высокое сходство авторов с хорошими статьями и, даже больше, является таковой, если цитируется хорошими статьями.*

На основе нашего анализа мы определили две особенности в первоначальной формулировке CИТЕХ. Они являются следующими:

1. $x_i^{(k)} \sim x_i^{(k-1)} \sum_{p=1}^n W_{ip}$: оценка автора CИТЕХ для автора i может расти по причине высокой продуктивности и, более того, если автор стремится выступать в соавторстве в небольших коллективах. Это позволяет в случае весьма продуктивного единичного автора превалять за счет алгоритма. Он (автор) даже может не нуждаться в росте популярности из-за подсчета ссылок (от хороших или иных авторов), чтобы получить высокую авторскую оценку CИТЕХ.

2. $y_j^{(k)} \sim y_j^{(k-1)} \sum_{p=1}^n W_{ip}$: оценка статьи CИТЕХ может просто вырасти при наличии повтора списка одинаковых авторов в значительной доле массива, с подобным явлением, становящимся более явным, если список сокращается. Также такие случаи могут превалять за счет CИТЕХ без роста подсчета ссылок (из хороших статей или в иных случаях).

Для иллюстрации возможных проблем, связанных с этими особенностями, сделаем два небольших вычисления, аналогичные предложенным Палом и Руджем [1]. Они показаны на рис. 1 и 2.

В результате особенностей, показанных на рис. 1 и 2, можно ожидать, что ранжирования автора и статьи, созданные CИТЕХ, будут испытывать на себе проблемы этих особенностей, поскольку экстремальные (чрезмерные) признаки публикации и ссылок соединятся. Задача данной статьи – предложить более элегантный вариант алгоритма CИТЕХ, изучающего упомянутые выше вопросы.

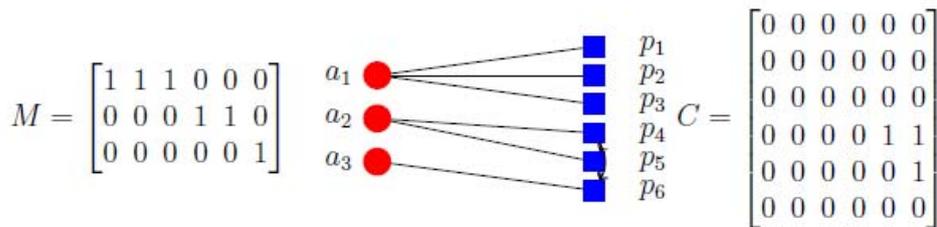


Рис. 1. Проблема 1 – гипотетический случай одного продуктивного единичного автора без ссылок.

СІТЕХ дает $x = [0,333; 0,333; 0,333]$ и $y = [0,143; 0,143; 0,143; 0,095; 0,191; 0,285]$.

Следовательно, все три автора ранжируются одинаково, даже если имеются резкие качественные различия между их моделями публикации и ссылок. Понятно, что статья p_6 имеет самую высокую оценку, следующую за p_5 , благодаря числу ссылок, которые они получают, по сравнению с отсутствием ссылок для других статей. Любопытно, что p_4 ранжируется ниже, чем статьи p_1, p_2 и p_3 , несмотря на нахождение под автором a_2 , который имеет на одну ссылку больше, чем a_1 (через p_5).

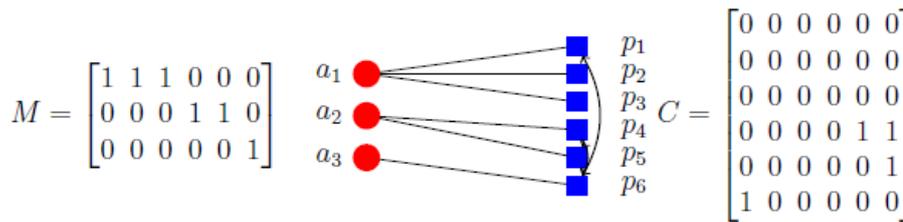


Рис. 2. Проблема 2 – эффект высокого сходства авторов с хорошими статьями.

Структура на этой диаграмме похожа на рис. 1 с одной дополнительной связью ссылок, добавленной из статьи p_6 к p_1 .

СІТЕХ дает $x = [0,521; 0,214; 0,214]$ и $y = [0,243; 0,175; 0,175; 0,068; 0,136; 0,203]$. Автор a_1 лидирует по авторской оценке, следующей за связью между a_2 и a_3 , несмотря на отсутствие самоцитирования (со)авторов на a_1 (отметим, что a_2 имеет самоцитирование одного автора через $p_4 \rightarrow p_5$). Статья p_1 ранжируется выше всех, несмотря на наличие только одной ссылки, поскольку она цитируется хорошей статьей (p_6). Благодаря способу оценки статей, пропагандируемому в СІТЕХ, статьи p_2 и p_3 также получают высокие оценки просто за счет наличия высокого сходства авторов со статьей p_1 .

УЛУЧШЕННЫЙ СДВОЕННЫЙ АЛГОРИТМ ОЦЕНКИ АВТОР-СТАТЬЯ

Как представлено в предыдущем разделе, СІТЕХ имеет встроенную тенденцию производить порядковый ранг, который дает нежелательный приоритет высокопродуктивным авторам (даже если они относительно мало влиятельны), помимо присвоения высокой относительной важности статьям, связанным с высокопродуктивными авторами (аннулируя влияние ссылок других статей).

Для обхода этих проблем мы предлагаем опустить член самоцитирования $(I+C^T)$ в уравнениях (1) и (2) и заменить матрицы M матрицами W , чтобы обеспечить сохранение подсчета ссылок при движении от сети ссылок статьи к сети ссылок автора (межавторские ссылки дробятся). Это приводит к следующему ряду равенств, определяющих наш **сдвоенный алгоритм оценки автор-статья**:

$$x^{(k)} = WC^T W^T x^{(k-1)} \quad (10)$$

$$y^{(k)} = C^T W^T x^{(k)} \quad (11)$$

Следуя предыдущим конвенциям [17,1], начинаем с вектора исходного предположения (конкретно, $x(0) = 1_{m \times 1}$ и $y(0) = 1_{n \times 1}$) и определяем значения оценок итеративно (т.е. повторное $k \geq 1$ до тех пор пока не достигается конвергенция для заданного допустимого уровня).

Уравнение (10) определяет критерий того, что «хороший автор цитируется хорошими авторами». Уравнение (11) устанавливает критерий того, что «хорошая статья цити-

руется хорошими авторами». Приведенные выше уравнения представляют логичную основу для повторного улучшения [18, с. 355-359]. Это можно увидеть, записав, что $L = WC$:

$$x^{(k)} = WL^T x^{(k-1)} = Wy^{(k-1)} \quad (12)$$

$$y^{(k)} = L^T x^{(k)} \quad (13)$$

Следовательно, *хороший автор имеет хорошие статьи, которые цитируются хорошими авторами, имеющими хорошие статьи* и т.д. Матрица размера $m \times n$, L , имеет записи $(L)_{ij} = \sum_{p=1}^n W_{ip} C_{pj}$, которые соответствуют кумулятивным дробным ссылкам, сделанным цитирующим автором i через статьи p (если принадлежат авторству i) на некоторую цитируемую статью j . По сути L кодирует *матрицу ссылок автор-статья*.

Записи произведения WL^T матрицы размера $m \times m$ в уравнении (12) отвечают кумулятивным дробным ссылкам, полученным авторами в строке i от авторов в столбце a . Это происходит, потому что $(WL^T)_{ia} = \sum_{p_1, p_2=1}^n W_{ap_2} C_{p_2 p_1} W_{ip_1}$ означает, что автор a в статье p_2 цитирует статью p_1 , которая принадлежит (со)авторству i . Сумма всех возможных статей p_1 служит объединению всех дробных ссылок, полученных автором i от автора a . WL^T , таким образом, является (дробной) *матрицей ссылок автора*.

В действительности, в связи с этим мы получаем, что авторская оценка, определенная в уравнении (12), соответствует $x_i^{(k)} = \sum_{a=1}^m \sum_{p_1, p_2=1}^n W_{ap_2} C_{p_2 p_1} W_{ip_1} x_a^{(k-1)}$. Следовательно, авторская оценка для автора i пропорциональна кумулятивным авторским ссылкам, полученным наравне с оценкой цитирующих авторов. Это приводит к интуитивной догадке, что *авторы продвигают друг друга через свои опубликованные работы*. Подобным образом, уравнение (13) предполагает, что оценка статьи для статьи j составит $y_j^{(k)} = \sum_{i=1}^m L_{ij} x_i^{(k)}$. Это определяет отношение, в котором *относительная важность статьи зависит от авторитета ее цитирующих авторов*.

ЭМПИРИЧЕСКАЯ ПРОВЕРКА

Мы проверяем CITEХ и CAPS (алгоритм дуальной оценки автор-статья) на статьях, опубликованных под предметной категорией «Информатика и библиотековедение» в Указателе цитированных журналов по науке и технике Института научной информации Томсона с 1980-х гг. по 2012 г. Этот массив данных содержит 213 530 статей 471 191 межстатейную ссылку и 73 597 авторских ключевых слов. Нами не проводятся поиски несоответствий в авторах и библиографических ссылках для того, чтобы изучить качество результата алгоритмов двоянной оценки автор-статья и CITEХ при использовании минимальной предварительной обработки данных.

Авторы

Результат схемы ранжирования зависит от того, как она оценивает выбранные свойства, представленные (или отсутствующие) для каждого данных относительно оставшейся части массива данных. В целом трудно определить эффективность соответствующего алгоритма оценки, когда нет справедливой основы для обоснования таких оценок. В подобных этому случаях наиболее чувствительным аспектом для реализации служит обращение к свойствам этих оценок, порожденных интересующим алгоритмом, а также выяснение вопроса относительно того, показывают ли созданные ранжирования разумное согласие с известными методами и наблюдениями.

В этом отношении распределение авторских оценок для алгоритма двоянной оценки автор-статья и CITEХ проявляет значительно высокий коэффициент корреляции ранга Спирмана (ρ) с оценкой h -индекса ($\rho < 0,01$): особенно, 0,77 и 0,69 для двоянной оценки автор-статья и CITEХ, соответственно. h -индекс [19] обеспечивает полезное сравнение с CAPS и CITEХ, так как он излишне сочетает особенности публикаций и ссылок. Тем не менее, в отличие от двоянной оценки автор-статья (и в меньшей степени от CITEХ) h -индекс не предназначен для проведения различий относительно того, получена ли ссылка из сравнительно «хорошей» статьи (автора) или иным образом, а значит, следует ожидать некоторого разночтения (несогласия) в итоговом ранжировании, что отражено в табл. 1.

Таблица 1

25 (из 73 597) верхних авторов по подсчету публикаций, подсчету ссылок, авторской оценке CAPS, авторской оценке CITEХ и h -индекса, соответственно. Значения h -индекса в колонках, обозначенных h , подсчитаны с использованием доступных данных (только статьи Института научной информации, индексированные в категории «Информатика и библиотековедение» Указателя цитированных журналов по науке и технике за 1980-2012 гг.) Отметим использование порядкового ранжирования для колонки h -индекса.

Ранг	Публикации		Количество раз цитирований		CAPS		CITEХ		h -индекс	
	b	Основной автор	b	Основной автор	b	Основной автор	b	Основной автор	b	Основной автор
1	5	Rogers M.	5	Davis F. D.	21	Egghe L.	5	Rogers M.	30	Glanzel W.
2	0	Cassada J.	29	Benbasat I.	26	Leyderdorff L.	0	Cassada J.	29	Bates D.W.
3	0	Klett R. E.	12	Venkatesh V.	23	Rousseau R.	0	Klett R. E.	29	Benbasat I.
4	1	Ramsdell K.	29	Bates D.W.	30	Glanzel W.	1	Ramsdell K.	28	Garfield E.
5	1	Christian G.	1	Pawlak Z.	24	Thelwall M.	1	Christian G.	26	Leydesdorff L.
6	0	Vicarel J.A.	19	Straub D.W.	16	Burrell Q.L.	0	Vicarel J.A.	25	Schubert A.
7	2	Hoffert B.	30	Glanzel W.	25	Schubert A.	2	Hoffert B.	25	Spink A.
8	1	Sutton J.	1	Gruber. T. R.	17	Ingwersen P.	1	Sutton J.	24	Grover V.
9	1	Sutton J. C.	25	Spink A.	17	Bar-ilan J.	24	Thelwall M.	24	Moed H.F.
10	1	Bigelow D.	14	Salton G.	22	Braun T.	21	Egghe L.	24	Thelwall M.
11	1	Stevens N.	3	Furnas G.W.	18	Cronin B.	30	Glanzel W.	23	Rousseau R.
12	0	Zlendich J.	24	Grover V.	17	Van.raan A..F.H.	26	Leydesdorff.L.	22	Braun T.
13	0	Faichild C.A.	3	Deerwester S.	16	White H.D.	1	Sutton J.C.	21	Egghe L.
14	1	Pearl N.	3	Dumais S. T.	12	Jasco P.	2	Decandido G.A.	20	Willett P.
15	0	Richard O.	2	Landauer T.K.	24	Moed H.F.	23	Rousseau R.	19	Ford N.
16	2	Gordon R.S.	8	Buckley C.	15	Vinkler P.	3	Stlifer E.	19	Saracevic. T.
17	0	Maccann D.	25	Schubert A.	16	Small H.	1	Bigelow D.	19	Smaglik P.
18	0	Lombardo D.	5	Morris M.G.	18	Mccain K.W.	25	Schubert A.	19	Straub D.W.
19	1	Williamson G.A.	1	Harshman R.	16	Bornmann L.	2	Rawlinson N.	18	Bates M.J.
20	5	Butler T.	7	Todd P.A.	28	Garfield E.	0	Davidson A.	18	Chen H.C.
21	1	Raiteri S.	9	Karahanna E.	9	Pao M.L.	0	De.baron F.H.K.	18	Cronin B.
22	1	Gillespie T.	26	Leydesdorff L.	15	Vaughan L.	0	Elizabeth P.	18	Dennis A.R.
23	1	Campbell P.	18	Zmud R.W.	6	Rao I.K.R.	1	Furlong C.W.	18	Lyytinen K.
24	3	Burns A.	14	Gefen D.	15	Daniel H.D.	0	Hammett D.	18	Mccain K.W.
25	2	Wyatt N.	19	Saracevic T.	15	Oppenheim C.	0	Hemingway H.	18	Zmud R.W.

Так как CAPS и CITECH также положительно соотносятся с $p=0,85$ ($p<0,01$), можно ожидать, что распределение h -индекса для верхних N рангов по оценке CAPS и CITECH напоминает друг друга с возрастанием N . Для верхних $N=25$ рангов, $\mu\text{CAPS}(h)=18,44$, тогда как $\mu\text{CITECH}(h)=6,76$. Для $N=2500$ средние значения h -индекса составляют 8,03 и 7,03, а для $N=2500$ получаем 3,32 и 3,36 для CAPS и CITECH соответственно. В идеале верхний процентиль любого ранжирования должен соответствовать легко интерпретируемой классификации по качеству, значит в этом смысле CAPS улучшает на авторском ранжировании CITECH (так как верхние ранги стремятся соответствовать высоким значениям h -индекса).

К слову сказать, автор верхнего ранга по CITECH (Rogers – с оценкой $3,37 \times 10^{-5}$) соответствует 83,6% целого распределения авторской оценки CITECH. Вместе с Cassada (авторская оценка = $2,39 \times 10^{-14}$) оба автора покрывают шокирующие 96% всех оценок. По всему списку авторов это соответствует коэффициенту Гини*, составляющему 0,9999. Для сравнения, 20% (14 719) верхних оцениваемых авторов согласно алгоритму CAPS насчитывают примерно 99,96% оценок (соответствуя коэффициенту Гини – 0,9891). Это предполагает, что различие между авторскими оценками CAPS для смежных рангов становится прогрессивно меньше при движении вниз по рангам. Это в большей степени преувеличено в CITECH.

Интересно, что коэффициенты Гини для дробного подсчета публикаций и дробного подсчета ссылок авторов в массиве данных по информатике и библиотековедению составляют 0,7744 и 0,8715 соответственно. Более того, 20% верхних авторов насчитывают 81,4% всех дробных публикаций наравне с 90% всех дробных ссылок. Хотя эти значения характерны для высоких уровней неравенств, они банально вполне сравнимы с уровнем неравенства, примененным CAPS. Наличие таких резких уровней неравенства предполагает широкий дифференциал в возможности исследователей области информатики и библиотековедения капитализировать ресурсы, технические навыки и возможности, имеющиеся в их распоряжении [20].

Статьи

25 верхних ранжирований путем подсчета ссылок, оценки статей CAPS и оценки статей CITECH даны в табл. 2. Самые верхние ранги CITECH представлены обменом статей одного и того же высокого по оценке автора (Rogers). Просматривая ранги, расположенные ниже 25 верхних, отмечаем, что за исключением статей с рангами 7-12, первые 3819 позиций занимают статьи под авторством Rogers, тогда как следующие 2610 позиций (ранги 3820-6429) принадлежат статьям автора Cassada. Значит, CITECH стремится превышать важность статей одного и того же высокооцененного автора, даже если они не отвечают работам «высокого влияния» или

* Коэффициент Гини является мерой статистической дисперсии, обычно используемой для измерения уровня неравенства в определенной выборке. Для выборки размера n , упорядоченной как $x_{i-1} \leq x_i$, он задается $G = \frac{2 \sum_{i=1}^n i x_i}{n \sum_{i=1}^n x_i} - \frac{n+1}{n}$. Коэффициент Гини, равный 1, показывает максимальное неравенство, в котором общая оценка связана только с одним элементом в выборке, тогда как оставшаяся часть выборки ничего не дает для общей оценки. Коэффициент Гини, равный 0, показывает идеальное равенство, где общая оценка распределяется одинаково между всеми элементами в выборке.

работам, воздействующим на «высоковлиятельные работы». Точно такой же эффект описан в разделе «Ожидаемое поведение и неясности».

Как видно в случае этих авторов, данные пристатейных ссылок показывают высокое неравенство, так как верхние 10% цитируемых статей насчитывают около 88,8% всех ссылок. Это ожидаемо, поскольку только доля статей цитируется и каждая из этих статей получает дополнительные внутренние связи ссылок в объеме, пропорциональном их актуальному числу внутренних связей ссылок. Это предполагает, что распределение ссылок управляется процессом кумулятивного преимущества/предпочтительного присоединения, в котором *богатый становится богаче* [6,21].

Для сравнения, оценка статей CAPS обладает коэффициентом Гини в 0,9912, тогда как CITECH имеет несколько более низкое значение – 0,9785. Это предполагает, что оба метода демонстрируют большие дифференциалы оценок между самыми верхними рангами. Что касается оценки статей CAPS, она может быть прослежена до того факта, что 81,2% самой низкой совокупности оценок имеет нулевую оценку (76% статей в данных исследования имеют ноль ссылок*). Причиной служит то, что соединение обеих авторских и статейных характеристик накладывает строгие ограничения на размер совокупности, имеющей не нулевые оценки. С другой стороны, оценка статей CITECH не имеет совокупности нулевых оценок (благодаря наличию искусственных пристатейных самоцитирований). Чрезвычайно высокие коэффициенты Гини для обеих оценок CITECH и CAPS предполагают, что мы можем обоснованно дифференцировать только малую долю массива данных, соответствующую верхним оценкам статей, которая согласуется с верхними оцениваемыми авторами.

Беглый взгляд на верхние оцениваемые статьи, приведенные в списке в колонке «Подсчет ссылок» табл. 2, показывает, что они в большей степени соответствуют статьям по информатике, а не информетрии. Противоположность этому составляет список, отраженный в колонке «CAPS», где внимание больше направлено в сторону статей по информетрии. Причина состоит в том, что алгоритм CAPS учитывает характеристики авторства при оценке статей, которые не фиксируются в простом подсчете ссылок. Так как авторы информетрии приведены в табл. 1, из этого следует, что статьи по информетрии также даны в табл. 2. Табл. 3 представляет список журналов в 100 верхних рангах. Она указывает на некоторую исследовательскую область, в преобладающей степени охарактеризованную каждым методом.

* Массив данных области информатики и библиотековедения содержит 103 768 статей из журнала *Library Journal* (~ 48,6 % от общего числа). Это примерно в 14 раз больше, чем во 2-м крупнейшем носителе, т.е. в журнале *Scientist*. Хотя это кажется чрезвычайно высоким, учтем, что только 1,9% статей из *Library Journal* дает 1% нулевых ссылок в массиве данных области информатики и библиотековедения (из общего числа 471 191 ссылка для 213 530 статей). Для сравнения, *Scientometrics* является только шестым крупнейшим вкладчиком в массив данных с 3100 статьями (1,5% всех статей в области информатики и библиотековедения), однако дает общее число в 29 792 ссылки (6,3% от общего числа области информатики и библиотековедения), становясь четвертым крупнейшим вкладчиком с точки зрения ссылок. Для справки: крупнейшие подсчеты ссылок принадлежат *MIS Quarterly*, *J AM MED INFORM ASSN* и *J AM SOC INFORM SCI* с 55 736, 30 470, 30 317 ссылками соответственно.

25 верхних (из 21 3530) статей по подсчету ссылок, оценке статей CAPS и оценке статей CИТЕХ. Статьи определяются по году публикации, согласно аббревиатуре журнального источника, тому, номеру страницы и первому автору. Аббревиатуры журнальных источников приведены в табл. 3. ТС обозначает число раз цитирований статей, как сообщается Институтом научной информации в 2012 г.

Коэффициенты корреляции ранга Спирмана ($p < 0,01$) для всех статей составляют:

$$p(C_1, C_2) = 0,87, p(C_1, C_3) = 0,17 \text{ и } p(C_2, C_3) = 0,26.$$

CAPS проявляется в лучшем совпадении с подсчетом ссылок, чем CИТЕХ.

Ранг	Подсчет ссылок (C_1)		CAPS (C_2)		CИТЕХ (C_3)	
	Статья	ТС	Статья	ТС	Статья	ТС
1	1982/IJCIS/11/341/pawlak	3319	2006/SCI/69/121/egghe	105	1995/LJ/120/113/rogers	1
2	1989/MISQ/13/319/davis	3251	1990/JIS/16/17/egghe	61	1995/LJ/120/119/rogers	1
3	1993/KA/5/199/gruber	2618	2006/SCI/69/131/egghe	250	1995/LJ/120/130/rogers	1
4	1990/JASIS/41/391/deerwester	2150	1998/JD/54/236/ingwersen	199	1995/LJ/120/187/rogers	1
5	1980/PAL/14/130/porter	1653	2005/S/19/8/braun	86	1995/LJ/120/213/rogers	1
6	2003/MISQ/27/425/venkatesh	1534	2003/JASIST/54/550/ahlgren	123	1996/LJ/121/100/rogers	1
7	1988/IPM/24/513/salton	1449	2006/SCI/69//169/braun	127	2011/LJ/136/30/fox	0
8	2001/MISQ/25/107/alavi	1075	2006/SCI/67/491/van.raan	177	2007/LJ/132/36/albanese	4
9	1995/ISR/6/144/taylor	1021	1999/JD/55/577/smith	93	1993/LJ/118/32/berry	2
10	2003/JMIS/19/9/delone	772	2001/JASIST/52/1157/thelwall	94	1989/LJ/114/18/decandido	1
11	2004/MISQ/28/75/hevner	724	1985/JD/41/173/egghe	48	1989/LJ/114/57/decandido	0
12	1995/MISQ/19/189/compeau	684	1992/IPM/28/201/egghe	41	1995/LJ/120/12/stlifer	1
13	2003/MISQ/27/51/gefen	677	1989/SCI/16/3/Schubert	165	1992/LJ/117/52/rogers	0
14	2000/ISR/11/342/venkatesh	596	1997/JD/53/404/almind	163	2000/LJ/125/91/rogers	0
15	1999/MISQ/23/67/klein	569	2001/SCI/50/65/bjorneborne	93	2005/LJ/130/172/rogers	0
16	2000/MISQ/24/169/bharadwaj	568	1986/SCI/9/281/Schubert	162	2006/LJ/131/114/rogers	0
17	1992/MISQ/16/227/adams	542	2006/SCI/67/315/glanzel	88	2006/LJ/131/114/rogers	0
18	1995/MISQ/19/213/goodhue	540	2006/SCI/69/161/banks	60	2006/LJ/131/123/rogers	0
19	1987/MISQ/11/369/benbasat	526	1996/SCI/36/97/egghe	31	2006/LJ/131/123/rogers	1
20	1999/MISQ/23/183/karahanna	513	1991/JASIS/42/479/egghe	29	2007/LJ/132/132/rogers	1
21	1999/JAMIA/6/313/bates	497	2003/SCI/56/357/glanzel	82	2007/LJ/132/132/rogers	1
22	1988/MISQ/12/259/doll	477	1986/SCI/9/103/leydsedorff	46	2007/LJ/132/171/rogers	0
23	1999/IJGIS/13/143/stockwell	475	2001/SCI/50/7/bar-ilan	64	2007/LJ/132/96/rogers	0
24	2000/MISQ/24/115/venkatesh	475	2002/JASIST/53/995/thelwall	72	2006/LJ/131/27/rogers	1
25	2003/ISR/14/189/chin	472	1996/JIS/22/165/egghe	24	2003/LJ/128/40/rogers	2

Таблица 3

Перечень журналов в 100 верхних рангах, составленный с помощью алгоритма

Подсчет ссылок			CAPS			CИТЕХ		
MISQ	mis.quart	43	SCI	scientometrics	33	LJ	Libr.j	100
ISR	inform.syst.res	14	JASIS	j.am.soc.inform.sci	17			
JAMIA	j.am.med.inform.assn	9	JASIST	j.am.soc.inf.sci.tec	14			
JASIS	j.am.soc.inform.sci	6	JD	j.doc	12			
JD	j.doc	3	JIS	j.inf.sci	9			
IPM	inform.process.manag	3	IPM	inform.process.manag	6			
JMIS	j.manage.inform.syst	3	JI	j.informetr	5			
IJCIS	int.j.comput.inf.sci	2	ARIS	annu.rev.inform.sci	2			
IM	inform.manage	2	SSI	soc.sci.inform	1			
IJGIS	int.j.georg.inf.sci	2	S	scientist	1			
SCI	Scientometrics	2						
ARIS	annu.rev.inform.sci	1						
CJLS	can.j.inform.sci	1						
EJLS	eur.j.inform.syst	1						
GIQ	gov.inform.q	1						
IJGIS	int.georg.inf.syst	1						
IMA	inform.manage-amster	1						
JASIST	j.am.soc.inf.sci.tec	1						
JIS	j.inf.sci	1						
KA	knowl.acquis	1						
OR	Online.rev	1						
PAL	Program-atom.lib	1						

ЗАКЛЮЧЕНИЕ

В этой статье мы построили модифицированную версию алгоритма CITEХ, первоначально введенного Палом и Руджем [1]. Алгоритм был разработан для присвоения оценок относительной важности статьям и авторам с учетом данных из обоих наборов одновременно. Традиционные методы, такие как подсчет ссылок и PageRank не могут, например, делать это без соответствующей модификации. Модификация CITEХ, которую мы предлагаем, это адаптированный алгоритм CAPS (своенная оценка автор-статья), разработанный с целью изучить некоторую слабость исходного алгоритма Пала и Руджа, которая описана в разделе «Ожидаемое поведение и неясности» (по сути, недостатки могут быть прослежены вплоть до искусственно введенных самоцитирований на уровне статьи).

Используя реальный массив данных (статьи Института научной информации, опубликованные в 1980-2012 гг. в Указателе цитированных журналов по науке и технике в предметной категории «Информатика и библиотковедение»), мы показываем, что наши предложенные модификации превосходят CITEХ в определении важных авторов и статей. Тем не менее, алгоритм CAPS кажется страдает от высокого неравенства в распределениях итоговой оценки, как показывается чрезвычайно высоким коэффициентом Гини (~ 0,99). Данное неравенство также подчеркивается в CITEХ. Это предполагает, что и CAPS, и CITEХ создают чрезмерное предубеждение в распределении оценок относительно меньшинства верхних оценок.

Однако, это совсем не обязательно плохо. По своей структуре CAPS распределяет высокие оценки авторам и статьям, связанным с примерами, где вероятность будущего успеха (рост в подсчете публикаций или подсчете ссылок) пропорциональна предыдущему успеху. Значит, CAPS можно использовать для выдвижения на первый план частей данных, принадлежащих эффекту «богатый становится богаче». Наоборот, CITEХ поощряет высокие оценки для авторов, находящихся в хвосте распределения публикационной активности, и по аналогии поощряет высокие оценки для статей, опубликованных такими авторами независимо от относительной важности их статей внутри сети цитирования. В этом смысле CITEХ полезен для поиска примеров, где высокая продуктивность плохо соотносится с низким влиянием.

Хотя библиометрические аналитические алгоритмы, такие как CITEХ или CAPS, или даже библиометрические адаптации алгоритмов ранжирования сайтов, такие как HITS или PageRank могут оказаться полезными в определении того, что важно в заданном массиве данных, необходимо осознавать ограничения и особенности таких методов. Каждый метод находит точно то, что он призван искать, и поскольку это трудно учитывать, отбросим каждую релевантную особенность или случайность, мы должны признать, что произведенные ранжирования сами являются только фасетами соответствующей организации данных. Значит, библиометрические аналитические алгоритмы должны использоваться в первую очередь для управления глобальными решениями (т.е. для выработки рекомендательных механизмов) и, если необходимо, применяться с крайней осторожностью при составлении умозаключений по относительному положению библиометрических объектов.

ЛИТЕРАТУРА

1. *Pal A., Ruj S.* CITEХ: A new citation index to measure the relative importance of authors and papers in

scientific publications. — 2015. — arXiv preprint arXiv:1501.04894.

2. *Bhatt A., Martens B.* The topics of CAAD: An evolutionary perspective/ Tidafi T. and Dorta T., editors, *Joining Languages, Cultures and Visions: CAAD Futures 2009// Proceedings of the 13th International CAAD Futures Conference, Montreal. Les Presses de l'Universite de Montreal, 2009.*

3. *Retblefsen M.L., Aldrich A.M.* Environmental health citation patterns: Mapping the literature 2008–2010// *Journal of the Medical Library Association: JMLA.* — 2013.— Vol. 101, No.1.— P.47.

4. *Simon H.* On a class of skew distribution functions. — *Biometrika*, 1955.—p. 425–440.

5. *de Solla Price D. J.* Networks of scientific papers// *Science.*— 1965.— Vol. 149, No. 3683. — P.510–515.

6. *de Solla Price D.* A general theory of bibliometric and other cumulative advantage processes// *Journal of the American Society for Information Science.* — 1976. — Vol. 27, No.5. — P.292–306.

7. *Newman M.* The first-mover advantage in scientific publication// *EPL (Europhysics Letters).*— 2009.— Vol. 86, No.6.— P. 68001.

8. *Chen P., Xie H., Maslov S., Redner S.* Finding scientific gems with Google's PageRank algorithm// *Journal of Informetrics.*— 2007.—Vol. 1, No.1. — P. 8–15.

9. *Franceschet M.* Pagerank: Standing on the shoulders of giants// *Communications of the ACM.*— 2011.— Vol. 54, No.6.— P. 92–101.

10. *Perron O.* Zur theorie der matrices// *Mathematische Annalen.* — 1907. — Vol. 64, No.2.— P. 248–263.

11. *Frobenius G. F.* Uber Matrizen aus nicht negativen Elementen. *Konigliche —Akademie der Wissenschaften*, 1912. — p. 456–477.

12. *Zhou D., Orshanskij S.A., Zha H., Giles C.L.* Co-ranking authors and documents in a heterogeneous network/ *Seventh IEEE International Conference on Data Mining, 2007(ICDM 2007)*, pages 739– 744. — IEEE, 2007.

13. *Brin S., Page L.* The anatomy of a large-scale hypertextual web search engine// *Computer Networks and ISDN Systems.* — 1998. — Vol. 30, No.1. — P. 107–117.

14. *Langville A.N., Meyer C.D.* *Google's PageRank and Beyond: The Science of Search Engine Rankings.*— Princeton University Press, New Jersey, USA, 2006.

15. *Jabne B.* *Computer vision and applications: A guide for students and practitioners.* — Academic Press, San Diego, CA, USA, 2000.

16. *Balakin K. V.* *Pharmaceutical data mining: Approaches and applications for drug discovery.* — John Wiley & Sons, Hoboken, New Jersey, USA, 2010.

17. *Kleinberg J.M.* Authoritative sources in a hyperlinked environment// *Journal of the ACM(JACM).*— 2010.— Vol. 46, No.5.— P. 604–632.

18. *Easley D., Kleinberg J.* *Networks, crowds, and markets.* —Cambridge Univ Press, 2010.

19. *Hirsch J.E.* An index to quantify an individual's scientific research output// *Proceedings of the National Academy of Sciences of the United States of America.* — 2005. — Vol. 102, No.46. — P.16569.

20. *Shockley W.* On the statistics of individual variations of productivity in research laboratories// *Proceedings of the IRE.* — 1957. — Vol. 45, No.3. — P. 279–290.

21. *Barabasi A., Albert R., Jeong H.* Mean-field theory for scalefree random networks// *Physica A: Statistical Mechanics and its Applications.* — 1999. — Vol. 272, No.1. — P.173–187.

Авторский указатель к МФИ, 2016, т. 41

	№	Стр.		№	Стр.		№	Стр.
Брини К.	3	19	Зилински Л.	3	19	Оливейра Э. Ф. Т.	2	32
Гарсиа-Эрнандес П.	1	15	Кальдерон-Реэчо А.	1	15	Пош Л.	2	43
Гобен А.	3	19	Коста М. П. да	4	3	Прагап Г.	4	20
Гонсалес-Фернандес-Виллависенсио Н.	1	15	Лейте Ф. С. Л.	4	3	Рабельо Р.	1	26
			Лоусон С.	1	3	Ратнавелу К.	4	20
Грасио М. К. К.	2	32		3	3	Сандерс К.	1	3
Грей Д.	3	3	Маури М.	3	3	Смит Л.	1	3
Джана К.	4	12	Мейерник М. С.	3	34	Филлипс Д.	3	34
Джонсон К. А.	2	3	Мигель С.	2	32	Франнсен Т. Ф.	2	16
Домингес-Ароса М.	1	15	Ниенхаус Э.	3	34	Юджум Э. А.	4	20
Дутта Б.	4	12	Николаисен Й.	2	16			

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 14.12.2016 г.

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

Усл. печ. л 4,00 Уч.-изд. л. 4,41 Тираж 33 экз.

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

Тел. (499) 155-44-95

Центр (Отдел) научно-информационного обслуживания (ЦНИО) ВИНИТИ РАН

Информационные услуги, предоставляемые ЦНИО ВИНИТИ РАН:

- проведение тематического поиска и консультации поисковых экспертов;
- подготовка списков научной литературы;
- подбор, копирование полнотекстовых материалов из первоисточников на бумажном носителе и в электронном виде;
- библиометрическая оценка публикационной активности исследователей и научных организаций с использованием российских и зарубежных баз данных;
- информационное обеспечение информационно-аналитической деятельности по подготовке и предоставлению аналитических обзоров и других научных материалов.

ВИНИТИ РАН располагает следующими информационными ресурсами:

- фондом НТЛ, включающим более 2,5 млн. отечественных и иностранных журналов, книг, депонированных рукописей, авторефератов диссертаций и другой научной литературы, ретроспектива – с 1991 года;
- базами данных и Интернет-ресурсами: БД ВИНИТИ (разработка ВИНИТИ), БД SCOPUS, БД Questel (патенты) и другими реферативными ресурсами;
- полнотекстовыми электронными ресурсами (статьи, патенты, материалы конференций).

Ознакомиться с информацией о доступных полнотекстовых и реферативных ресурсах можно на сайте ВИНИТИ www.viniti.ru

К услугам пользователей – **Электронный Каталог ВИНИТИ** <http://catalog.viniti.ru>
и **служба электронной доставки документов.**

Осуществляется платное информационное обслуживание по разовым заказам и на договорной основе с предоставлением всех необходимых финансовых документов.

Проводится индивидуальное обслуживание пользователей в читальном зале ЦНИО ВИНИТИ.

Обращаться в ЦНИО ВИНИТИ:

- адрес: 125190, Россия, г. Москва, ул. Усиевича, 20;
- телефоны: 8(499) 155 -42 -43, 8(499) 155 -42 -17;
- эл. почта cnio@viniti.ru, fdk@viniti.ru;
- факс 8(499) 930 -60 -00 (для ЦНИО).

База данных (БД) ВИНИТИ РАН

Федеральная база отечественных и зарубежных публикаций по естественным, точным и техническим наукам, генерируется с 1981 г., обновляется ежемесячно, пополнение составляет около 1 млн документов в год. Тематическое наполнение соответствует реферативному журналу ВИНИТИ. Для поиска одновременно по всем или нескольким тематическим фрагментам генерируется единая Политематическая БД.

БД ВИНИТИ РАН в сети INTERNET

Сервер ВИНИТИ - <http://www.viniti.ru> – обеспечивает on-line доступ к Базе данных ВИНИТИ РАН круглосуточно без выходных.

На основе БД ВИНИТИ РАН предоставляются следующие услуги:

- Диалоговый поиск научно-технической информации **в режиме on-line**;
- **Демо-версия**, позволяющая ознакомиться с основными функциями поисковой системы, составом данных, формами представления документов и получить навыки работы с системой;
- **Поисковые эксперты ВИНИТИ** выполняют тематический поиск по разовым или постоянным запросам, а также окажут **консультационные услуги**.

БД ВИНИТИ РАН на CD-ROM

Любые наборы тематических фрагментов БД ВИНИТИ или их разделов за любой период с 1981 г., а также **проблемно-ориентированные выборки** из БД ВИНИТИ по актуальным направлениям научных исследований могут быть предоставлены на договорной основе **в поисковой системе (ИПС) "Сокол"**, работающей под управлением Microsoft Windows и обеспечивающей следующие возможности:

- **Чтение** документов в режиме последовательного просмотра или выборочно по оглавлению за весь период заказанной ретроспективы
- **Поиск** документов по автору, заглавию, источнику, ключевым словам или словосочетаниям, реферату, рубрикам, году издания, стране, языку и т.д. (всего более 20 признаков)
- **Словарь** системы поможет правильно подобрать термины для поиска и выбрать глубину их усечения.
- Для **уточнения поиска** можно дополнительно использовать год издания документа, язык текста документа, рубрики, шифры тематических разделов БД.
- Выполненные **запросы можно сохранять** для их последующего использования и/или редактирования.

125190, г. Москва, ул. Усиевича, 20, БД ВИНИТИ РАН.

Отдел взаимодействия с потребителями – (499) 155-45-25, (499) 152-58-81

E-mail: csbd@viniti.ru, sales@viniti.ru

WWW: <http://www.viniti.ru>

УВАЖАЕМЫЕ КОЛЛЕГИ!

ВИНИТИ РАН предлагает Вашему вниманию Реферативный Журнал в электронной форме

РЖ в электронной форме (ЭлРЖ) выпускается по всем разделам естественных, технических и точных наук.

Каждый номер ЭлРЖ является полным аналогом печатного номера РЖ по составу описаний документов, их оформлению и расположению. Он сопровождается оглавлением, указателями.

ЭлРЖ представляет собой информационную систему, снабженную поисковым аппаратом и позволяющую пользователю на персональном компьютере:

- читать номер РЖ, последовательно листая рефераты;
- просматривать рефераты отдельных разделов по оглавлению;
- обращаться к рефератам по указателям авторов, источников, ключевых слов;
- проводить поиск документов по словам и словосочетаниям;
- выводить текст описаний документов во внешний файл.

ЭлРЖ в версии Windows Вы можете получить за текущий год с любого номера, а также за предыдущие годы.

Подробную информацию Вы можете получить:

Адрес: 125190, Россия, Москва, ул. Усиевича, 20, ВИНТИ РАН

Коммерческое управление

Телефон/Факс: 8 (499) 155-45-25, 8 (499) 152-58-81

E-mail: contact@viniti.ru, sales@viniti.ru