

СОДЕРЖАНИЕ

<b>Дутта Б.</b> Путешествие от либраметрии к альтметрии: взгляд в прошлое	3
<b>Чжу Я., Сяон Ю.</b> Определение науки о данных	13
<b>Брук М., Мюррей-Раст П., Оппенхайм Ч.</b> Социальные, политические и правовые аспекты извлечения текстов и данных	18
<b>Шоудари С. Р., Сингх А. К., Нелаканти А.</b> Ответ на поиск: предложение использовать поисковую информацию для лучшего представления контента сетевых сайтов	23
<b>Мацевичюте Е.</b> Отчет о международной конференции «Направления в издательстве и контексты 2014 г. – цифровые авторы и электронные книги», 8-9 декабря 2014 г., г. Пула, Хорватия	28
<b>Маир Ф., Шарнхорст А., Мутшке П., Фромхольц И.</b> Библиометрически улучшенный информационный поиск: 2-й международный семинар BIR	32

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Академик РАН **Ю.М. Арский** (Российская Федерация) — *главный редактор*,  
ВИНИТИ РАН, 125190, Москва, ул. Усневича, 20. Телекс 411249

Проф. д-р. **Р.С. Гиляревский** (Российская Федерация) — *заместитель главного редактора*,  
ВИНИТИ РАН, 125190, Москва, ул. Усневича, 20. Телекс 411249

**С. Дж. Паркер** (Канада) — *заместитель главного редактора*, IDCR, P.O. Box 8500,  
Ottawa, Ontario K1G 3H9, Canada

**А. Джикарайст** (Великобритания) — CURA Consortium and GAVEL g.e.i.e,  
38 Ship Street, Brighton BN1 1AB, UK

**М. Дрейк** (США) — Технологический институт шт. Джорджия, Библиотечный  
и информационный центр, 704 Cherry Street, Atlanta, Georgia 30332-0900, USA

**А. де Кемп** (Германия) — Издательство “Springer-Verlag”, Postfach 10 52 80,  
D-69042 Heidelberg, Germany

Д-р **Т. Кеннон** (Великобритания) — Отдел исследований и разработок  
Британской библиотеки, 2 Sheraton Street, London W1V 4BH, UK

**М. Миддлтон** (Австралия) — Школа информационных систем, QUT Gardens  
Point Campus, 2 George Street, Brisbane, 4000 QLD., Australia

**Т. Молвиг** (Норвегия) — Национальное управление по научной информации,  
вузовским и специальным библиотекам, P.O. Box 2439 Solli, N-0201, Oslo,  
Norway

**Х. Ринкон Феррейра** (Бразилия) — Бразильский институт информации по  
науке и технике (IBICT), SAS — Quadra 5, Lote 06, Bloco H, 700-70-000 Brasilia  
D.F., Brazil

**С. Феррейро** (Чили) — Чилийский университет, Системы информационных  
и библиотечных служб, Casilla de Correo 10D, Santiago, Chile

Проф. **Ю. Фуздивара** (Япония) — Университет Цукуба, Институт электроники  
и информатики, Tsukuba-shu, Ibaraki, 305 Japan

Д-р **М. Хименес** (Испания) — Испанское общество по научной документации  
и информации, Fuencarral, 123-6° dcha., 28010, Madrid, Spain

# Путешествие от либраметрии к альтметрии: взгляд в прошлое\*

Бидьярти ДУТТА  
(Bidyarthi DUTTA)

Отделение библиотекосведения  
и информатики, Университет  
им. Видьясагара, Индия

*Двадцатый век можно охарактеризовать как столетие развития метрических наук, то есть либраметрии, наукометрии, библиометрии, информетрии, эконометрии, технометрии, биометрии, социометрии, психометрии и т. д. ... Возможно, возникновение киберметрии было заключительной вехой продвижения метрических наук в прошлом столетии. Интернет и революция открытого доступа достигли пика в 1990-ые гг., завершив этот этап в 2001г. Рассвет нового тысячелетия дал вспышку искре света по всему земному шару, которой являлась Википедия. Процесс распространения знания уничтожил еще одну баррикаду, чтобы гарантировать людям легкий доступ. Это было новое понятие, которое кардинально преобразовало мир научной коммуникации. Различные новые измерения добавились к процессам сбора, хранения, обработки, распространения и оценки информации. Измерения эффективности и влияния этих процессов постепенно систематизировали их в направлении большей объективности. Таким образом, новые метрические науки постоянно развиваются в различных предметных областях. Недавние разработки включают викиметрию, метрику на уровне статьи, альтметрию и т.д. В этой статье все метрики, обычно напрямую или косвенно связываемые с масштабом и контекстом библиотекосведения и информатики, обсуждаются вместе с некоторым кратким историческим экскурсом. Начиная от либраметрии Ранганатана, они исследуются вплоть до самых новых метрических исследований. Все метрики категоризованы в три класса на основе соответствующего времени возникновения, т. е. классические метрики, неоклассические метрики и современные метрики. Четыре метрики, т. е. либраметрия, библиометрия, наукометрия и информетрия, категоризованы под классическими метриками. Киберметрия относится к неоклассической метрике, тогда как оставшиеся другие, развитые в XXI столетии, признаны современными метриками.*

## Введение

Термин «метрический» происходил от латинского слова *metricus* [1] и французского *métrique* [2], что означает меру чего-то, или любое значение полученного количественного измерения или приближения. Впервые это слово было использовано в 1864 г. [2]. Оно охватывает ряд предметных областей, например, общую теорию относительности в физике, организацию сетей, математику, анализ программного обеспечения и т.д. Одно из самых известных использований этого слова найдено как единица измерения в различных предметных облас-

тях, а также в библиотекосведении и информатике. Любая система измерения неизбежно включает один трудный с методологической точки зрения вопрос, который является ничем иным, как статистикой. Профессор Р.С. Махаланобис в начале пятидесятых годов описал статистику как ключевую технологию, поскольку она необходима для всех видов социально-экономической деятельности и в связи с тем, что статистические методы используются во всех разработках и прогнозирующих исследованиях. Статистика применяется почти ко всем главным предметным областям в широких дисциплинах, таких как проектирование, медицина, сельское хозяйство, социология, поведенческие науки, когнитивистика и т.д. Статистика и родственные методы при применении внутри предметной области могут дать начало новому предмету, например, применение статисти-

\* Перевод Dutta B. The journey from librametry to altmetrics: A look back.– <http://eprints.rclis.org/23665/2/B-Dutta-JU-Golden-Jubilee-Paper.pdf>

ки к общественным наукам или социологии дает начало новому предмету, т. е. социометрии. Другими подобными примерами являются эконометрия, измерения в сфере образования (educametry), психометрия и т.д. В области библиотековедения и информатики применения количественных методов сегодня очень популярны. С точки зрения классифицирующего специалиста, предмет *статистика* или *количественная техника*, может, таким образом, рассматриваться также как самостоятельная идея отдельно от ее существования как главного предмета науки. Она также включена во вспомогательные таблицы основных схем классификации (DDC, UDC или CC) точно так же, как другие предметы, такие как история, философия, управление и т.д. Ранние работы в этой предметной области уходят к середине 1920-ых гг., и исследования, включающие применения количественных методов к библиотечной и библиографической работе, широко известны как *статистическая библиография*. Уиттинг [3] рассматривал вопрос об истории статистической библиографии, указывая, что происхождение термина могло привести к его первому использованию Колом и Илесом [4] в 1917 г., а также Халмом [5] в 1922 г. Термин “статистическая библиография” означает применение количественных методов к библиотечным и библиографическим работам. Вслед за Халмом Госнелл [6] также использовал этот термин в докторской диссертации “Темп устаревания в книжном собрании библиотеки колледжа”, которая была представлена Нью-Йоркскому университету в 1943 г. Он использовал этот термин, чтобы сделать акцент скорее на количественном аспекте, чем на качественных особенностях. Рейзиг [7] также утверждал, что имеется потенциальная польза статистической библиографии как метода анализа информационных потребностей в медицинских науках. Ключевое слово “статистическая библиография” сформировано из двух слов, т. е. статистики и библиографии. Слово *статистика* означает [8] “факты или данные цифрового типа, которые могут быть собраны, классифицированы и сведены в таблицу для представления информации о предмете”. Слово *библиография* [8] произошло от двух латинских слов, *biblion* и *graphos*, первое означает книги, а второе – писать. Слово *библиография* означает список работ авторов или литературы, касающийся определенного предмета или автора. Халм [5] определил библиографию как науку организации знания. Эти два слова *статистика* и *библиография* объединены, чтобы сформировать ключевое слово статистическая библиография, которая широко ассоциируется с применением количественных методов в библиотековедении.

### Классическая метрика

Ранганатан [9] уже в 1948 г. на Конференции Aslib в Лемингтоне предложил, “что библиотекарям необходимо развивать «либраметрию» наряду с биометрией, эконометрией и психометрией, так как многие вопросы, связанные с библиотечной работой и услугами, имеют дело с большими количествами”. Слово ‘librametry’ или ‘librametrics’ представляет собой сочетание двух отдельных слов, библиотека и метрика, которое указывает на применение математических моделей и статистических методов в оценке библиотечных систем и услуг. Хорошо известный факт, что любая библиотека, включающая или собрание библиотеки, или библиотечные услуги, может быть достаточно объективной только тогда, когда представлена через статистические методы. Так как статистика говорит точным языком, то в ней нет

места для каких-либо неясностей. Как прокомментировал Рао [9], “несмотря на раннюю попытку определить область либраметрии, предмета, едва развитого до начала 1970-ых гг.” и Причарда [10] использовал термин «библиометрия» в 1969 г., чтобы описать все исследования, стремящиеся измерить процесс письменной коммуникации. Он [10] определил термин следующим образом: “определение и цель библиометрии состоят в том, чтобы пролить свет на процесс письменной коммуникации, природу и направленность дисциплины (поскольку это проявляется через письменную коммуникацию) с помощью подсчета и анализа различных аспектов письменной коммуникации”. Появление термина библиометрия часто приписывается Причарду, но Уилсон [11] указал, что этот термин является французским прецедентом. Фонсека [12] указал на использование французского эквивалента термина (‘bibliometrie’) Полем Отле [13] в 1934 г. в его *Трактате о документации* (Traité de Documentation. Le livre sur le livre. Theorie et Pratique), который является мало известной работой, а Раздел 124, с.13-22 этого текста называется «Книга и мера. Библиометрия» (Le Livre et la Mesure. Bibliometrie). Сенгупта [14] утверждает, что Кемпбелл [15] произвел первое библиометрическое исследование, используя статистические методы для того, чтобы изучать тематическое рассеивание в публикациях. Глоссарий Британского стандарта терминов документации [16] описывал библиометрию “как исследование использования документов и модели публикации, в которой могли быть применены математические и статистические методы. Поттер [17] определил это как “исследование и измерение моделей публикации всех форм письменной коммуникации и их авторов”. Сенгупта [18] описал это как “организацию, классификацию и количественную оценку моделей публикации всей макро- и микрокоммуникации наряду с их авторством с помощью математического и статистического исчисления”. Герцель [19] определил это как “науку записанного дискурса, которая использует определенные методологии, математические и научные, в своем исследовании и является контролируемым исследованием коммуникации. Это совокупность литературы, библиография количественно, численно или статистически проанализированная – статистическая библиография; библиография, в которой измерения используются, чтобы документировать и объяснять регулярность феномена коммуникации”. Библиометрические исследования, как правило, делятся на два класса – описательную библиометрию и оценочную библиометрию. Описательные исследования касаются географического распределения или временного распределения подсчета продуктивности. Оценочная библиометрия включает ссылки или цитирования, которые известны как подсчет использования литературы. Согласно Николасу и Ритчи [20], “библиометрия предоставляет информацию о структуре знания и о том, как оно передается”. Они проводят четкое различие между двумя типами библиометрии, т. е. “той, которая описывает характеристики или особенности литературы (описательная) и той, которая исследует отношения, сформированные между компонентами литературы (поведенческая). Оценочные исследования используют ссылки на литературу, применяемую учеными области. Компетенция библиометрии включает изучение отношения в пределах литературы или описания литературы, сосредотачиваясь на последовательных моделях, включая авторов, монографии, журналы или предмет”. Фэрторн [21] в 1969 г. определил библиометрию как “количественную

обработку свойств записанного дискурса и относящегося к этому поведения”.

Согласно Рао [9], понятие библиометрии в Восточной Европе было известно как наукометрия в начале семидесятых годов. Худ и Уилсон [22] упоминали, что в 1969 г. Налимов В.В. и Мульченко З.М. [23] предложили русский эквивалент термина ‘scientometrics’ (наукометрия) [24]. Название этого термина очевидно и главным образом используется для исследования всех аспектов литературы по науке и технике. Этот термин достиг широкого признания, которому способствовало учреждение Гибором Брауном в Венгрии в 1978 г. журнала *Scientometrics*. В соответствии с его подзаголовком *Scientometrics* включает все количественные аспекты науковедения, коммуникации в науке и научной политики [25]. Цели и возможности области наукометрии были также отражены в некоторых ранних статьях Налимова, который помог в создании возникающей дисциплины [26,27]. Наукометрия, таким образом, охватывает несколько предметов, таких как количественные аспекты науки, научная политика, администрирование науки и т.д. Фактически предметная область наукометрии имеет дело с количественными исследованиями результатов всех научных дисциплин. Внутренний смысл термина *bibliometrics* (*библиометрия*) говорит о количественных исследованиях или статистическом анализе библиографии и ее различных использований. Возможности наукометрии, таким образом, шире, чем возможности библиометрии. Раджан и Сен [28] в общих чертах обрисовали возможности наукометрии – “этимологически наукометрия означает исследование, касающееся измерения в сфере науки. Наука, исходя из многих точек зрения, может измеряться по следующим аспектам: производство (выпуск) дипломированных специалистов, аспирантов или докторов науки; создание научных институтов, учреждений по исследованиям и обучению в сфере науки; развертывание научных трудовых ресурсов, утечка мозгов; расходы НИОКР; развитие средств научной коммуникации, например первичной и вторичной научной периодики; научная литературная и научно-информационная система, услуги и продукты. Метрические исследования всех этих аспектов находятся в пределах наукометрии. Область наукометрии, которая имеет дело с научной информацией, также охвачена информетрией. Следует отметить, что очень большая доля литературы по информетрии относится к наукометрии. Согласно Таге [29], “наукометрия - исследование количественных аспектов науки как дисциплины или экономической деятельности. Это часть социологии науки и имеет применение в разработке научной политики. Она включает количественные исследования видов научной деятельности, содержащих, помимо прочего, публикацию, и поэтому в какой-то степени перекрывает библиометрию”. Из основного определения ясно, что возможности наукометрии ограничены изучением науки, тогда как исследования информетрии распространяются на все области знания. Согласно Ланкастеру [30], предмет наукометрии вырос из простого анализа данных в хорошо определенный предмет, включающий прикладную статистику, моделирование, воспроизведение, кластерный анализ, исследование сети цитирования и т.д.

Слово *информация* встроено в термин информетрия, а информация, подобно воздуху, находится здесь, там и всюду, она вездесуща. В контексте любого предмета понятие информации может пониматься как фундаментальный стронтельный блок, поскольку, если нет информации, то никакое обучение, размышление или

процесс обмена/культивирования знания не может начаться. Встроенные единицы в предварительно обсужденных трех словах, т. е. либраметрии, библиометрии и наукометрии (*librametrics*, *bibliometrics* и *scientometrics*) – это библиотека, библиография и наука соответственно. Эти три слова не настолько фундаментальны, как информация. Первое слово указывает на учреждение, где информация доступна; второе указывает на объект, включающий список систематических записей некоторых объектов, несущих информацию, и последнее – на широкую дисциплину изучения и исследования. Непосредственное сочетание этих трех слов с информацией прямо говорит о происхождении или фундаментальности слова информетрия (*informetrics*) по сравнению с либраметрией, библиометрией или наукометрией (*librametrics*, *bibliometrics* или *scientometrics*). Термин информетрия был предложен в 1979 г. Отто Накке [29] из Западной Германии, в то время как три другие термина были придуманы в 1948 г. либраметрия (*librametrics*) и в 1969 г. библиометрия и наукометрия (*bibliometrics* и *scientometrics*). Хотя *информетрия* является более поздним, чем три другие термина, но логично принять его в контекстуальном и концептуальном плане более фундаментальным и идущим впереди *libra-*, *biblio-* и *sciento-* *metrics*. Информетрия может пониматься как наиболее обобщенный термин, который включен в категорию наукометрии, а конкретнее – библиометрии. Раджан и Сен [28] из INSDOC (теперь NISCAIR), Нью-Дели, обрисовали в 1985 г. цели информетрии, как обеспечение надежных данных для исследования и разработки; выработку тактики и планирования; измерения учреждений, проектов, программ и видов деятельности. Брукс [31] в 1989 г. указал, что это определение информетрии является самым широким и самым глубоким из трех метрических терминов, т. е. *biblio-*, *sciento-* и *info-* *metrics*. Кроме того, Рао [9] указал, что информетрия в основном используется для обозначения использования и развития разнообразия мер для исследования и анализа некоторых свойств информации вообще и документов, в частности. Краткое определение было дано Эгте и Руссо [32] в подзаголовке их книги: “Информетрия: количественные методы в библиотекведении, документации и информатике”. Информетрия покрывает эмпирические исследования литературы и документов, а также теоретические исследования других математических законов и свойств наряду с распределениями, которые были обнаружены. Таге-Сатклифф [29] определил этот термин следующим образом: “информетрия является исследованием количественных аспектов информации в любой форме, не только записей или библиографий, и в любой социальной группе, а не только ученых. Таким образом, она рассматривает как количественные аспекты неформальной или устной коммуникации, так и письменной, а также информационных потребностей и использований обычных людей, не только интеллектуальной элиты. Она (информетрия) может включать, использовать и расширять многие исследования по измерению информации, которые лежат за рамками библиометрии и наукометрии. ... Два явления, которые в прошлом не рассматривались как часть библиометрии или наукометрии, но вполне соотносятся с рамками информетрии, являются следующими: определение и измерение информации и типы и особенности измерения поисковой эффективности”. Ингверсен и Кристенсен [33] представили следующее определение: “термин информетрия определяет недавнее расширение традиционных библиометрических исследований также в целях охвата

ненаучных сообществ, в которых информация производится, передается и используется”. Уилсон [25] завершает последний обзор ARIST следующим определением: “...информетрия - количественное исследование собраний единиц среднего размера потенциально информативного текста, направленное на научное понимание процессов информирования на социальном уровне”. Раджан и Сен [28] выдвигают несколько новых точек зрения и указали на несоответствия определения Накке относительно информетрии. Накке определил информетрию как “применение математических методов к исследованию объектов информатики с целью описания и анализа их свойств и законов, чтобы оптимизировать эти объекты в принятии решений [34]”. Раджан и Сен [28] полагают, что “это определение неадекватно, так как оно рассматривает только исследование ‘объектов информатики’, а не ‘виды деятельности информатики’ и ‘непосредственно информации’”. Согласно этому определению, информетрия стремится описывать и анализировать свойства и законы объектов информатики. Здесь следует указать, что информетрия не только описывает и анализирует свойства и законы объектов информатики, но также устанавливает законы, использующие математические и статистические методы, относящиеся к росту, распространению, использованию и распаду информации. Моралес [35] определил информетрию как ‘метрическую дисциплину, касающуюся исследования математических и статистических методов и моделей, их применения к количественному анализу их структуры, свойств научной информации, моделей и законов процессов научной коммуникации, включающих идентификацию собственных законов’. Бонич [36] выяснил причины, стоящие за появлением этого термина. Он сравнил ‘информетрию’ с ‘библиометрией’ и ‘наукометрией’ и заметил, что введение этого нового термина было необходимо для дифференцирования центральных областей информетрии, т. е. научной коммуникации, науковедения и библиотекведения.

В 1984 г. Всесоюзный институт научно-технической информации (ВИНИТИ) учредил Комитет по информетрии Международной федерации по документации (МФА) под председательством О. Накке, где «информетрия» (‘informetrics’) рассматривалась как общий термин и для библиометрии и для наукометрии. Такое использование было принято в ВИНИТИ в монографии Горковой В.И. в 1988 г. с русским названием *Информетрия* [Informetrics]. На первой Международной конференции по *Библиометрии и теоретическим аспектам информационного поиска* в 1988 г. Брукс предположил, что ‘информетрия’, которая относится как к библиометрии, так и наукометрии, также может иметь будущее [22] для документальной и электронной информации. *Informetrics 87/88* было принято как краткое название для опубликованных трудов конференции. Редакторы отметили, что “при продвижении нового названия классическим способом является использовать новое название наряду со старым” [37]. На второй Конференции в 1990 г. Брукс [38] подтвердил ‘информетрию’ как общий термин и для наукометрии и для библиометрии, акцентировав, что наукометрия более склонна к исследованиям политики, а библиометрия – к библиотечным исследованиям. Термин ‘информетрия’ получил большую значимость на *Третьей Международной конференции по информетрии* [39]. На четвертой Конференции, названной *Международной конференцией по библиометрии, информетрии и наукометрии*, все три термина акцентировались в равной степени. Труды четвертой конференции были изданы в четырех отдельных томах, три из которых представля-

лись целыми выпусками регулярных журналов на английском языке [40,41,42]. Исторически эта конференция очень важна, так как на ней было основано Международное общество по наукометрии и информетрии (ISSI), последующие конференции [43,44,45] проходили каждые два года под эгидой общества. Таким образом, термин ‘информетрия’ был широко признан к началу 1990-х гг. [25].

Одним из стойких приверженцев в контексте классической метрической области был американский ученый Юджин Гарфилд, родившийся в 1925 г. в Нью-Йорке [46]. Он получил докторскую степень по структурной лингвистике в Университете шт. Пенсильвания в 1961 г. и учредил Институт научной информации (ISI) в Филадельфии (шт. Пенсильвания, США). ISI сегодня образует основную часть научного подразделения компании Thomson Reuters. Гарфилд ответственен за многие инновационные библиографические продукты, включая *Current Contents*, *Science Citation Index (SCI)* и другие базы данных цитирования, а также журнал *Journal Citation Reports* и *Index Chemicus*. Он – редактор-основатель и издатель журнала *The Scientist*, новостного журнала для ученых в области наук о жизни. В 2007 г. он запустил HistCite, пакет программного обеспечения по библиометрическому анализу и визуализации. Следуя идеям, инспирированным известной статьей Ванневара Буша “Как мы можем мыслить” (1945 г.), Гарфилд предпринял разработку исчерпывающего индекса цитирования, показывающего распространение научного мышления; он основал Институт научной информации в 1955 г. *SCI* дал возможность вычислять импакт фактор, индекс непосредственности, полураспад цитированной литературы, полураспад цитирующей литературы и многие другие показатели для измерения важности научных журналов. Это привело к неожиданному открытию, что несколько журналов, таких как *Nature* и *Science* были ядром для всех точных наук. Однако такая модель не встречается в случае гуманитарных и общественных наук.

### Киберметрия: неоклассическая метрика

Мир научной коммуникации стоял перед некоторыми радикальными переменами с начала 1990-х гг. Печатная коммуникация на основе бумаги должна была постепенно уйти в сторону, чтобы освободить место новичку-пришельцу – электронным средствам массовой информации на основе онлайн (а также автономной) коммуникации. Последнее десятилетие XX века можно рассматривать как открытые ворота к Интернету и революции в коммуникации, которая добавила новые измерения относительно информации и интеллектуального общества. Кибернетика является часто используемым термином для интернет- и онлайн-коммуникации, который был предложен Норбертом Винером [46] в 1948 г. в контексте “научного исследования контроля и коммуникации в животном и машинном мире”. До понятия Интернета тогда было еще далеко. Фактически слово кибер является широко используемым общим префиксным термином, и его первое использование в качестве приставки отмечалось в слове кибернетика. Сен [47] отмечал, что “Винер взял из греческого языка слово, означающее рулевого, в которое заложена идея относительно контроля. После этого Уильям Гибсон использовал слово в рассказе, опубликованном в журнале *Omni*. Еще раз он использовал это слово в романе *Neuromancer*, появившемся в 1984 г. В романе слово несло смысл электронного пространства. С годами значение слова расширилось и теперь охватывает информаци-

онную технологию, Интернет и виртуальный мир". Широко известно, что применение математических и статистических методов является сердцем и душой таких предметов, как библиометрия, наукометрия, информетрия и т.д. То же самое справедливо и для киберметрии (cybermetrics). Эта специфичная область тесно связана с библиометрией, информетрией и наукометрией. В киберметрии вебсайты играют ту же самую роль, что документы в библиометрии, информетрии или наукометрии. Появление Интернета и Всемирной паутины в 1990-х гг. вызвало возникновение множества понятий, таких как киберпространство, кибербезопасность, киберзаконы, киберпреступления и т.д. Печатные документы постепенно вымирают, чтобы освободить пространство электронным средствам. Библиометрические или информетрические исследования, как правило, основаны на печатных документах, тогда как киберметрические исследования – основаны на сети. Сен [47] описал возможности киберметрии следующим образом: "время от времени создаются новые вебсайты в киберпространстве. Некоторые из них весьма часто представляют динамические происходящие изменения, некоторые – устойчивые, мало изменяющиеся или совсем не изменяющиеся, а немногие иногда исчезают без какого-либо предупреждения. Многие обладают длинной жизнью, а некоторые – короткой. Как различные периодические издания по одному и тому же предмету отличаются по качеству, так и вебсайты по одной и той же теме отличаются по качеству. Развились и продолжают развиваться механизмы, чтобы оценить вебсайты, вычислить сетевые импакт факторы и изучить цитированные сайты. Все эти явления ставят вебсайты в рамки юрисдикции киберметрии".

Вебометрия (webometrics), синонимичный термин по отношению к киберметрии, также часто используется в контексте измерения ссылок, структуры и моделей использования Всемирной паутины. Бьернборн и Ингверсен [48] определили вебометрию как "исследование количественных аспектов строительства и использования информационных ресурсов, структур и технологий в сети, описываемое библиометрическими и информетрическими подходами". Термин вебометрия был сначала предложен Алмайндом и Ингверсеном [49] в 1997 г. Другое определение давалось Телволлом [50], т. е. "исследование основанное на сети контента прежде всего количественными методами для научных целей социологии, использующих методы, которые не являются определенными для какой-то одной области исследования". Это определение подчеркивает развитие прикладных методов для использования в более широких социальных контекстах. Оно расширило возможности вебометрии в отношении разнообразия других областей общественных наук вне сферы информатики и библиотекведения. Одна важная мера - "импакт фактор сети" (WIF), введенный Ингверсеном [51] в 1998 г. Его можно определить как число веб-страниц на сетевом сайте, получающих ссылки от других вебсайтов, разделенное на число веб-страниц, опубликованных на сайте и доступных краулеру. Сегодня также развивается множество других показателей для измерения влияния сети.

### **Анализ цитирования**

Согласно Гарфилду и Рубину, анализ цитирования может определяться как полное исследование частоты, моделей и графов ссылок в статьях и книгах [52,53]. Он использует ссылки в научных работах, чтобы установить

связи с другими научными работами. Анализ цитирования - один из наиболее широко используемых методов библиометрии. Например, библиографическое сочетание и социотирование являются ассоциативными мерами, основанными на анализе ссылок. Автоматизированное индексирование ссылок [54] добавило новое измерение к исследованию анализа ссылок, позволяя проанализировать миллионы ссылок для крупномасштабных моделей и открытия знания. Первым примером автоматизированного индексирования ссылок был CiteSeer, за ним следовал Google Scholar. Сегодня способы анализа цитирования легко доступны для вычисления различных мер влияния ученых, основанных на данных от указателей цитирования. Они имеют многочисленные применения, начиная от идентификации опытных рецензентов до обзорных статей и предложений по предоставлению грантов, до обеспечения прозрачных данных в поддержке качества научного рецензирования, срока пребывания в должности и решений о продвижении в карьере.

Первое, насколько известно, задокументированное исследование анализа цитирования, основанное на подсчете и анализе ссылок, было описано в 1927 г. благодаря Гроссу и Гроссу [55]. Они подсчитали и проанализировали ссылки к статьям в журнале по химии и путем ранжирования названий журналов в соответствии с числом полученных ссылок составили список журналов, которые рассматривали в качестве 'обязательных в химическом образовании'. Это первое зарегистрированное исследование анализа ссылок. Следует отметить, что исследования Коула и Илеса были основаны на записях в библиографиях, а не на ссылках. Анализ ссылок постепенно стал неотделимой областью библиометрии с применениями во многих областях знания.

### **Феномен рассеивания**

Библиографические компоненты, такие как авторы или источники (журналы, книги, монографии и т.д.) всегда рассеиваются в любой предметной области по большому количеству элементов. Например, статьи рассеяны по журналам, авторы – по журналам или статьям и т.д. Это явление известно как библиографическое рассеивание. Одна из главных целей классических метрических исследований состоит в измерении такого явления рассеивания. Массивы данных, обычно обрабатываемые в классических метрических исследованиях, скорее считаются строго подчиняющимися ряду законов и математических распределений. Три самых известных из таких законов - закон Брэдфорда, закон Лотки и закон Ципфа. Эти законы и их разнообразные математические формы рассматривались несколькими авторами, включая Фэрторна [56], Брукса [57], Лейкулера [58], Букстейна [59] и Прайса [60]. Некоторые авторы выражали эти законы в готовых для использования формах в случае практических ситуаций, в то время как другие исследовали сходства этих законов со стандартными статистическими распределениями. Некоторые ученые рассматривали эти три закона как особые случаи основного уникального распределения. В частности, Прайс [60] предложил распределение кумулятивного преимущества, статистически моделирующее ситуацию, в которой успех приносит успех. Прайс полагал, что это распределение служит "подходящей основной вероятностной теорией для закона Брэдфорда, закона Лотки, закона Парето, закона Ципфа и для всех эмпирических результатов анализа частоты цитирований".

## Современные метрики

### *Викиметрия (Wikimetrics): еще одна вежа*

Термин викиметрия – объединение слов *wiki* и *metrics*. До настоящего времени Википедия, возможно, является одним из самых захватывающих проектов лидеров XXI века. Джимми Уолес и Ларри Санджер положили начало Википедии 15 января 2001 г. и данное ей название представляло комтамминацию слов *wiki* (название типа совместного вебсайта, от гавайского слова "быстрый") и *энциклопедия*. Только спустя тринадцать лет, т. е. в феврале 2014 г., "Нью-Йорк Таймс" сообщила, что Википедия в глобальном масштабе считается пятой среди всех вебсайтов, утверждая, что "с просмотрами 18 миллиардов страниц и почти 500 миллионами уникальных посетителей в месяц ... Википедия отстает только от Yahoo, Facebook, Microsoft и Google, самого крупного игрока с 1,2 миллиардами уникальных посетителей [61]". Викиметрия облегчает анализ данных страниц Вики, устанавливая стандартные метрики вдоль движения и улучшая взаимодействие между заинтересованными в данных лицами. Викиметрия – сетевое средство (прежде известное как UserMetrics), разработанное для упрощения измерения локальной деятельности пользователя, основанное на ряде стандартизированных метрик. Используя этот инструмент, другие метрики могут быть выбраны и применены к произвольно определенной комбинации пользователей для оценки их общей производительности, пребывания, количества и качества работы вики. Платформа является лингво- и проектно-инвариантной (она может находить данные из любого проекта Викимедиа), расширяемой (с помощью добавления новой метрики, изменения метрических параметров) и предназначенной для создания массива данных для различных типов групп анализа и более дружелюбной по отношению к пользователю оценки программы [62].

### *Метрика открытого источника*

Метрическое программное обеспечение - мера некоторого свойства части программного обеспечения или его спецификаций. Так как количественные измерения существуют во всех науках, то есть непрерывное усилие со стороны практиков и теоретиков вычислительной науки в целях перенести подобные подходы на разработку программного обеспечения. Цель – получение объективных, воспроизводимых и количественно измеримых мер, которые могут иметь многочисленные ценные применения в составлении схем и планирования бюджета, оценке стоимости, проверке обеспечения качества, отладке программного обеспечения, оптимизации эффективности программного обеспечения и оптимальном назначении персонала для реализации задачи [63]. Важная функция метрического программного обеспечения была предпринята проектом FLOSSMetrics, где FLOSSMetrics означает Free/Libre Open Source Metrics. Главная цель FLOSSMETRICS – это строительство, издание и анализ крупномасштабной базы данных с информацией и метрикой о разработке бесплатного программного обеспечения, вытекающее из нескольких тысяч проектов по программному обеспечению, которые используют существующие методологии и уже разработанные средства. Слово *libre* – заимствованное в английском языке слово из французского языка, оно означает 'состояние свободы' или 'обладание свободой' или свобода. FLOSSMetrics предоставляет доступ к гуглам материалов этой базы данных (наряду со

схемами, таблицами и другой количественной информацией о проектах развития FLOSS) на вебсайте Melquiades. FLOSSMetrics также обеспечивает FLOSS Guide для SMEs (малых и средних предприятий), что объясняет выгоды FLOSS, формирующие SMEs, и как его использовать в своих интересах [64].

Основные цели FLOSSMetrics можно суммировать следующим образом:

- Идентифицировать и оценивать источники данных и развивать исчерпывающую структуру базы данных, построенную на результатах CALIBRE;
- Объединять уже доступные инструменты, чтобы извлекать и обрабатывать такие данные в полной платформе;
- Строить и поддерживать обновленную эмпирическую базу данных, применяющую инструменты извлечения к тысячам проектов открытых источников;
- Развивать методы визуализации и аналитические исследования, особенно касающиеся эталонного тестирования, идентификации лучших практик, измерения и прогнозирования успеха и неудачи проектов, измерения производительности, моделирования и оценки стоимости/усилия;
- Распространять результаты, включая данные, методы и программное обеспечение;
- Предусматривать эксплуатацию результатов, представляя план эксплуатации, утвержденный совместно с участниками проектов от промышленности, особенно с учетом точки зрения SME.

### **Метрика журнала**

Самая главная метрика журнала – *импакт фактор* – была введена Юджином Гарфилдом в 1975 г. для журналов, которые внесены в указатель Journal Citation Report. Она в высокой степени зависима от научной дисциплины, возможно, от скорости, с которой статьи цитируются в области. Процент всех цитирований, имеющих место в первые два года после публикации, широко варьируется среди дисциплин – от 1-3% в математических и физических науках до 5-8% в биологических науках [65]. Таким образом, логически вероятно, что импакт факторы не могут использоваться для сравнения журналов по дисциплинам. Импакт фактор основан на среднеарифметическом числе ссылок на статью, но обычно наблюдается, что подсчеты ссылок следуют распределению Брэдфорда или распределению степенного закона. Среднее арифметическое, таким образом, – статистически несоответствующая мера. Например, приблизительно 90% импакт фактора журнала *Nature* в 2004 г. были основаны только на четверти его публикаций, и, таким образом, важность любой публикации будет отличаться от общего числа [66]. Кроме того, сила отношений между импакт факторами журналов и темпами цитирования статей в них постоянно уменьшается, так как статьи стали доступны в цифровом формате [67]. Сегодня имеется множество других метрик, чтобы измерить качество или стандарт журналов. Они перечислены ниже:

1. **Индекс непосредственности (Immediacy Index):** определяется как отношение числа статей, процитированных в один год, к числу статей, изданных в том же году.
2. **Цитированный полуураспад (Cited Half Life):** определяется как средний возраст процитированных статей.



3. **Совокупный импакт фактор (Aggregate Impact Factor):** определяется как импакт фактор для всего предмета.

4. **Оценка собственного фактора (Eigenfactor Score)** [68,69]: оценка собственного фактора – рейтинг общей важности научного журнала. Она разработана Джевином Уэстом и Карлом Бергстромом в университете Вашингтона в 2008 г. Журналы оцениваются согласно числу поступающих ссылок, со взвешенными ссылками из высокоранжируемых журналов, а не из низкоранжируемых, чтобы сделать больший вклад в собственный фактор. Как мера важности, оценка собственного фактора растет с общим влиянием журнала. Журналы, производящие более высокое влияние в области, имеют большие оценки собственного фактора. Это может быть определено как число раз, которое статьи, опубликованные в журнале за 5 лет, цитировались в году. Здесь ссылки на тот же самый журнал удалены. Бергстром и Уэст [68,69] также интерпретировали этот показатель как модифицированный 5-летний импакт фактор. Что касается года JCR, то алгоритм собственного фактора эффективно ранжирует журналы согласно ссылкам и отрезку времени, в течение которого ученые находятся на вебсайте журнала. Это, по сути, исследование ссылки вебсайта журнала.

5. **Оценка влияния статьи (Article Influence Score):** оценка влияния статьи предоставляет измерения относительной важности журнала на постатейной основе. Это оценка собственного фактора журнала, деленная на долю статей, опубликованных журналом. Эта доля нормализуется так, чтобы общее количество статей (сумма) из всех журналов было равно 1. Средняя оценка влияния статьи равняется 1,00. Оценка большая, чем 1,00 показывает, что каждая статья в журнале имеет влияние выше среднего. Оценка меньше 1,00 указывает, что каждая статья в журнале имеет влияние ниже среднего. Что касается года в JCR, то влияние статьи журнала ISI определяется как “оценка собственного фактора (Eigenfactor), деленная на долю всех статей ISI, опубликованных журналом ISI”. В 2006 г., самым важным журналом по оценке влияния статьи был журнал *Annual Reviews of Immunology*, с влиянием статьи – 27,454. Это означает, что средняя статья в этом журнале имеет в двадцать семь раз больше влияния, чем средний журнал в JCR [70].

6. **h5-индекс:** определяется как наибольшее число  $h$ , такое, что, по крайней мере,  $h$  статей в этой публикации были процитированы, не менее  $h$  раз. Журнал с  $h5$ -индексом, равным 8, имеет за 5 лет, по крайней мере, 8 статей, которые были процитированы 8 раз.

7. **Показатель SJR – SCImago Journal Rang:** определяется как число ссылок, полученных числом статей, взвешенных в зависимости от престижа и предметной области журнала. Он измеряет научное влияние академических журналов, которое учитывает и число ссылок, полученных журналом, и важность или престиж журналов, являющихся источником таких ссылок.

8. **SNIP - нормализованное источником влияние на статью (Source-Normalized Impact Per Paper):** SNIP измеряет вытекающее из контекста влияние ссылки источника, взвешивая ссылки, основанные на общем количестве ссылок в предметной области. Оно помогает сделать непосредственное сравнение источников в различных предметных областях. SNIP принимает во внимание особенности предметной области источника, которым является набор документов, цитирующих этот источник. SNIP главным образом рассматривает:

- частоту, с которой авторы цитируют другие статьи в своих библиографических списках;
- скорость, с которой наступает влияние ссылки;
- степень, до которой используемая в оценке база данных охватывает литературу области.

SNIP – это отношение среднего подсчета цитирования источника на статью и потенциал цитирования ее предметной области. Потенциалом цитирования предметной области источника является среднее число ссылок на документ, цитирующий этот источник. Это представляет вероятность для документов быть процитированными в определенной области. Источник в области с высоким потенциалом цитирования стремится иметь высокое влияние на статью. Потенциал цитирования важен, поскольку он принимает во внимание тот факт, что типичные подсчеты цитирования широко варьируются между научными дисциплинами. Например, они имеют тенденцию быть выше в науках о жизни, чем в математике или общественных науках. Если статьи в одной предметной области содержат в среднем 40 процитированных ссылок, в то время как в другой области – 10, тогда первая область имеет потенциал цитирования, который в 4 раза выше, чем во второй. Потенциал цитирования также изменяется между предметными областями в пределах дисциплины. Например, основные журналы имеют тенденцию показывать более высокие потенциалы цитирования, чем прикладные или медицинские журналы, а журналы, охватывающие новые появляющиеся темы, обладают тенденцией иметь более высокие потенциалы цитирования, чем периодические издания во вполне устоявшихся областях [71].

### Метрика автора

Метрика автора по существу измеряет научное влияние академических публикаций соответствующих авторов. Все авторы имеют ограничения. Метрика должна подчеркнуть точный параметр для измерения. Поскольку различные предметные области или дисциплины имеют различные типы моделей публикации, то научное влияние авторов, принадлежащих ко всем дисциплинам, не может быть измерено в равном стандарте. Согласно Кауру [72], анализ влияния автора все больше играет важную роль в оценке предоставления гранта, решениях о найме и сроках пребывания в должности. Главной целью данной метрики являются, однако, оценки научной и социальной перспективы авторов. Основные параметры метрики автора перечисляются ниже [73]:

1. **Ссылки:** число раз, процитированное в литературе.

2. **Использование:** число раз просмотров на вебсайте (издателей); число раз загрузок; как часто получался доступ к дополнительным данным.

3. **Сборы данных:** как часто это было отмечено в кладаках/ обменено (CiteULike/Mendeley).

4. **Упоминания:** количество размещений в блогах; как много новых историй; упоминания в Википедии и т.д.; комментарии к вебсайту издателей и в другом месте.

5. **Социальные СМИ:** обмен/ лайки в Facebook; доли LinkedIn; Твиты.

6. **h-индекс:** ученый имеет индекс  $h$ , если  $h$  из его статей  $N_p$  имеют, по крайней мере,  $h$  ссылок каждая, а другие  $(N_p-h)$  статьи имеют не больше, чем  $h$  ссылок каждая. Этот индекс был введен Дж. Е. Хиршем в 2005 г.

7. **g-индекс:** учитывая набор статей, ранжированных в порядке убывания числа полученных ими ссылок, g-индекс – это уникальное наибольшее число такое, что

самые важные  $g$  статьи получили (вместе), по крайней мере,  $g^2$  ссылок. Индекс был введен Лео Эгге в 2006 г.

8. **i-10 индекс:** i10-индекс показывает число академических публикаций, написанных автором, которые имеют, по крайней мере, десять ссылок из других работ. Этот индекс был введен в июле 2011 г. Google как часть их работы над Google Scholar.

9. **h-ядро:** h-ядро публикации – это множество ведущих процитированных  $h$  статей из публикации. Они являются статьями, на которых  $h$ -индекс основан.

10. **h-медиана:**  $h$ -медиана публикации – медиана подсчетов ссылок в ее  $h$ -ядре.

11. **h5-индекс, h5-ядро и h5-медиана:** этими показателями публикации являются, соответственно,  $h$ -индекс,  $h$ -ядро и  $h$ -медиана только тех ее статей, которые были изданы в последние пять полных календарных лет.

Web of Knowledge сталкивается со следующими параметрами оценки автора:

1. Найденные результаты.
2. Сумма раз цитирования.
3. Сумма раз цитирования без самоцитирований.
4. Цитирующие статьи.
5. Цитирующие статьи без самоцитирований.
6. Среднее число ссылок на документ.
7.  $h$ -индекс.

## Метрики уровня статьи

Метрики уровня статьи – это метрики для измерения использования и влияния индивидуальных научных статей. Традиционно использование и влияние исследования оценивались на основе библиометрии или информетрии, которые обычно сосредотачивались на журналах. Сюда относятся импакт фактор или индекс непосредственности.

Некоторые метрики уровня исследователя, такие как  $h$ -индекс,  $g$ -индекс или  $i-10$  индекс развились в последнем десятилетии. Метрика уровня статьи в отличие от метрики журнала или метрики автора не сосредотачивалась на журналах или авторах, а концентрировалась на отдельной статье. Это связано с альтметрией, но отличается от нее. Издательство открытого доступа PLOS обеспечивает метрики уровня статьи для всех своих журналов, включая загрузки, ссылки и альтметрию.

## Альтметрия

Альтметрия – это новая метрика, предложенная как альтернатива широко используемому импакт фактору журнала и личным индексам ссылок, таким как  $h$ -индекс,  $g$ -индекс или  $i-10$  индекс. Термин альтметрия был введен Джейсоном Примом [74,75] в 2010 г. как обобщение метрики уровня статьи и внедрен в хештег `twitter#altmetrics`. Хотя об альтметрии часто думают как о метрике, относящейся к статьям, она может применяться и к людям, журналам, книгам, массивам данных, презентациям, видео, хранилищам исходного текста, сетевым страницам и т.д. Альтметрия охватывает не только подсчеты ссылок, но также и другие аспекты влияния работы, такие как, сколько баз данных и знания относятся к ней, виды статей, загрузка или упоминания в социальных СМИ и новостных медиа. Альтметрия – очень широкая группа метрик, охватывающая различные части влияния, которое может иметь статья или работа. Классификация альтметрии была предложена ImpactStory в сентябре 2012 г. и очень похожая классификация используется PLOS [76,77].

## Заключение

В этой статье прослеживается краткая история всех метрик в контексте библиотекведения и информатики. Обсуждение начато с либраметрии, самой передовой метрики в библиотекведении, и продолжено до современности, т. е. альтметрии, которая является сегодня темой активного исследования. Все метрики распределены по трем категориям на основе соответствующего времени их введения, т. е. классическая метрика, неоклассическая метрика и современная метрика. Четыре метрики, то есть либраметрия, библиометрия, наукометрия и информетрия, категоризированы под классической метрикой. Киберметрия рассматривается как неоклассическая метрика, в то время как оставшиеся другие, разработанные в XXI веке, признаны современными метриками. Непрерывное концептуальное продвижение и необходимые поправки в идеях, таких как альтметрия, метрика уровня статьи, социометрия и т.д., продолжаются в целях достижения новых результатов в областях измерения эффективности и анализа влияния исследования и научной коммуникации.

## Литература

1. *Metric*. — 2014// Wikipedia.— <http://en.wiktionary.org/wiki/metric>.
2. *Metrics*. — 2014// Merriam-Webster Online: Dictionary and Thesaurus.— <http://www.merriam-webster.com/dictionary/metric>.
3. *Wittig G. R.* Statistical bibliography - a historical footnote// *Journal of Documentation*.—1978.— Vol. 34, No.3.— P. 240-241.
4. *Cole F. J., Eales N. B.* The history of comparative anatomy. Part 1: A statistical analysis of the literature// *Science Progress*.— 1917.—Vol. 11.— P. 578-596.
5. *Hulme E. W.* Statistical bibliography in relation to the growth of modern civilization. — London: Grafton, 1923.
6. *Gosnell C. F.* The rate of obsolescence in college library book collection. (Ph.D dissertation).— New York University, New York City, USA,1943.—P. 16.
7. *Raisig L. M.* Statistical bibliography in health sciences// *Bulletin Medical Libraries Association*.— 1962.— Vol.50.— P. 450.
8. *Friend J. H., Guralnik D. B. (Ed.)*. Webster's New World Dictionary of American Language.— New York: The World Pub Co, 1964.—P. 144.
9. *Rao I. K. R.* Informetrics: Scope, definition, methodology and conceptual questions// *Workshop on Informetrics and Scientometrics*, 16-19 March, Bangalore.— Documentation Research and Training Centre, Indian Statistical Institute, 1998.
10. *Pritchard A.* Statistical bibliography and bibliometrics// *Journal of Documentation*.—1969.— Vol. 25, No.4.— P. 348- 349.
11. *Wilson C. S.* The formation of subject literature collections for bibliometric analysis: The case of the topic of Bradford's Law of Scattering.( Ph.D. dissertation). — The University of New South Wales, Sydney, Australia, 1995. — <http://www.library.unsw.edu.au/~thesis/adt-NUN/public/adt-NUN1999.0056>.
12. *Fonseca E. N. D. A.* Bibliografia Estatística e Bibliometria: Uma Reivindicacao de Prioridades. [Statistical bibliography and bibliometrics: A re-indication of priorities]//*Ciencia da Informacao*. — 1973.— Vol. 2, No. 1.— P. 5–7.
13. *Otlet P.* Traite de Documentation. Le Livre sur le Livre. Theorie et Pratique. [Treatise on documentation. The

book on the book: Theory and practice]. — Brussels: Van Keerberghen, 1934.

14. *Sengupta I. N.* Bibliometrics, informetrics, scientometrics and librametrics: An overview// *Libri*.— 1992.— Vol. 42.— P. 75–98.

15. *Campbell F. B. F.* The theory of the national and international bibliography: With special reference to the introduction of system in the record of modern literature.— London: Library Bureau, 1896.

16. *British Standard Institution.* British standard glossary of documentation terms// Prepared under the direction of Documentation Standards Committee.— 1916.— Vol. 7.

17. *Potter W. G.* Introduction to bibliometrics// *Library trends*.— 1981.— Vol. 30, No 2.— P. 5-7.

18. *Sengupta I. N.* Bibliometrics, a bird's eye view// *IASLIC Bulletin*. — 1985.— Vol. 30.— P. 167-174.

19. *Hertzfel D. L.* Bibliometrics, history of the development of ideas in// *Encyclopaedia of Library and Information Science*.— 1987.— Vol. 42, No. 7.— P. 144-219.

20. *Nicholas D., Ritchie M.* Literature and bibliometrics.— London: Clive Bingley, 1978.— P. 9-11.

21. *Fairthorne R. A.* Empirical hyperbolic distributions (Bradford-Zipf-Mandelbort) for bibliometric description and predication// *Journal of Documentation*.—1969.— Vol. 25, No. 4.— P. 319-343.

22. *Hood W. W., Wilson C. S.* The literature of bibliometrics, scientometrics and informetrics// *Scientometrics*.— 2001.—Vol. 52, No. 2.—P. 291-314.

23. *Nalimov V. V., Mulchenko Z. M.* Eshche raz k voprosu o kontseptsii eksponentsial'nogo rosta. [A word to add on the exponential growth concept.] // *Nauchno-Tekhnicheskaya Informatsiya. Seriya 2*. — 1969.— No 8.— P. 12–14. [English translation// *Automatic Documentation and Mathematical Linguistics*.— 1969.— No. 3.— P. 37–40.]

24. *Nalimov V. V., Mulchenko Z. M.* Naukometriya. Izuchenie razvitiya nauki kak informatsionnogo protsessa. [Scientometrics study of the development of science as an information process].— Nauka, Moscow, 1969. (English translation.— Washington, D.C.: Foreign Technology Division. U.S. Air Force Systems Command, Wright-Patterson AFB, Ohio. — 1971. (NTIS Report No.AD735-634).

25. *Wilson C. S.* Informetrics/ M. E. Williams (Ed.)// *Annual Review of Information Science and Technology*, Vol.34 (pp. 3-143). — Medford, NJ: Information Today Inc. for the American Society for Information Science, 2001.

26. *Nalimov V. V.* Influence of mathematic statistics and cybernetics on the methodology of scientific investigations// *Zavodskaya Laboratoriya*.— 1970.— Vol. 36, No. 10.— P. 1218–1226. [English translation // *Industrial Laboratory*.— Vol. 36, No. 10.— P. 1549–1558.]

27. *Nalimov V. V., Kordon I. V., Korneeva A. Y. A.* Geograficheskoe raspredelenie nauchnoi informatsii. [Geographic Distribution of Scientific Information.]// *Informatsionnye Materialy*. Moscow: an SSSR Nauchnyi Sovet po Kompleksnoi Probleme Kibernetiki. [Informational Papers. Moscow: Soviet Academy of Science, Scientific Council on Cybernetics.] — 1971.— Vol. 2.— P. 3–37. [English translation// V. V. Nalimov. *Faces of Science*. Philadelphia, Institute for Scientific Information.— 1981.— Chapter 11.— P. 237–260.]

28. *Rajan T. N., Sen B. K.* An essay on informetrics: A study in growth and development// *Annals of Library Science and Documentation*. — 1986.— Vo. 33, No.1-2.— P. 1-12.

29. *Tague-Sutcliffe J. M.* An introduction to informetrics// *Information Processing and Management*. — 1992.— Vol. 28, No. 1.— P. 1-3.

30. *Lancaster, F. W.* Bibliometric method in assessing productivity and impact of research// Sarada Ranganathan Endowment for Library Science, Bangalore. — 1991.— P. 52.

31. *Brookes B. C.* Biblio-, sciento-, infor-metrics??? What are we talking about? /L. Egghe & R. Rousseau (Eds). *Informetrics 89/90*. Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics (pp. 31-43). —Amsterdam: Elsevier, 1990.

32. *Egghe L., Rousseau R.* Introduction to informetrics: Quantitative methods in library, documentation and information science.— Amsterdam: Elsevier Science Pub, 1990.

33. *Ingversen P. F., Christensen H.* Data set isolation for bibliometric online analyses of research publications: Fundamental methodological issues// *Journal of the American Society for Information Science*. — 1997.— Vol. 48.— P. 205–217.

34. *Nacke O.* Informetric: Ein nuer name fuer eine disziplin// *Nachrichten fuer Dokumentation*. — 1979.— Vol. 30, No.6.— P. 219-226.

35. *Morales M.* Informetrics and its importance// *International Forum on Information and Documentation*.— 1985.— Vol. 10, No.2.— P. 15-21.

36. *Bonitz M.* Scientometrie, bibliometrie, informetrie// *Zentralblatt für Bibliothekswesen*.— 1982.— Vol. 96, No. 2.— P. 19-24.

37. *Egghe L., Rousseau R.* (Eds.) *Informetrics 87/88: Select Proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*; 1987 August 25-28; Diepenbeek, Belgium. — Amsterdam: Elsevier, 1988.

38. *Egghe L., Rousseau R.* (Eds.) *Informetrics 89/90: Selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics*; 1989 July 5-7; London, Ontario, Canada. — Amsterdam: Elsevier, 1990.

39. *Rao I. K. R.* (Ed.) *Informetrics - 91*. Selected papers from the Third International Conference on Informetrics; 1991 August 9-12; Bangalore, India. — Bangalore: Sarada Ranganathan Endowment for Library Science, 1992.

40. *Glänzel W., Kretschmer H.* (Eds.) Selected papers presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; 1993 September 11-15; Berlin, Germany// *Research Evaluation*. —1992.— Vol. 2, No. 3.— P. 121-188.

41. *Glänzel W., Kretschmer, H* (Eds.) Selected papers presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; 1993 September 11-15; Berlin, Germany// *Scientometrics*.— 1994.— Vol. 30, No.1.

42. *Glänzel W., Kretschmer H.* (Eds.) Selected papers presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics; 1993 September 11-15; Berlin, Germany// *Science and Science of Science*.— 1994.— Vol. 3, No. 5.

43. *Koenig M. E. D., Bookstein A.* (Eds.) Fifth Biennial Conference of the International Society for Scientometrics and Informetrics; 1995 June 7-10; River Forest, IL. — Medford, NJ: Learned Information Inc., 1995.

44. *Peritz B. C., Egghe L.* (Eds.) Sixth Conference of the International Society for Scientometrics and Informetrics; 1997 June 16-19; Jerusalem, Israel. — Jerusalem: The School of Library, Archive and Information Studies of the Hebrew University of Jerusalem, 1997.

45. *Macias-chapula C. A.* (Ed.) Seventh Conference of the International Society for Scientometrics and Informetrics;

- 1999 July 5-8; Colima, México. — Colima: Universidad de Colima, 1999.
46. *Cybernetics*. — 2014// Wikipedia.— <http://en.wikipedia.org/wiki/Cybernetics>.
47. *Sen B. K.* Cybermetrics-meaning, definition, scope and constituents//Annals of Library and Information Studies. — 2004.— Vol. 51, No. 3.— P. 116-120.
48. *Björneborn L., Ingwersen P.* Toward a basic framework for webometrics// Journal of the American Society for Information Science and Technology. — 2004.— Vol. 55, No. 14.— P. 1216–1227.
49. *Almind T. C., Ingwersen P.* Informetric analyses on the World Wide Web: Methodological approaches to 'webometrics'// Journal of Documentation.—1997.— Vol. 53, No. 4.— P. 404–426.
50. *Thehwall M.* Introduction to webometrics: Quantitative web research for the social sciences.— California: Morgan & Claypool, 2009.
51. *Ingwersen P.* The calculation of web impact factors// Journal of Documentation.— 1998.— Vol. 54, No. 2.— P. 236–243.
52. *Garfield E.* Citation indexing - its theory and application in science, technology and humanities. — Philadelphia: ISI Press, 1983.
53. *Rubin R.* Foundations of library and information science (3rd ed.). — New York: Neal-Schuman Pub, 2010.
54. *Giles C. L., Bollacker K. D., Lawrence S.* CiteSeer: An automatic citation indexing system// Digital libraries 98: The Third ACM Conference on Digital Libraries, June 23–26, 1998, Pittsburgh, PA. —New York: Association for Computing Machinery, 1998.— P. 89–98.
55. *Gross P. L. K., Gross E. M.* College libraries and chemical education// Science.— 1927.— Vol. 66.— P. 1229- 1234.
56. *Fairthorne R. A.* Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction// Journal of Documentation.— 1969.— Vol. 25.— P. 319-343.
57. *Brookes B. C.* Bradford's law and the bibliography of science// Nature. — 1969.— Vol. 224, No. 5223.— P. 953-956.
58. *Leimkuhler F. F.* The Bradford distribution// Journal of Documentation.—1967.— Vol. 23, No. 3.— P. 197-207.
59. *Bookstein A.* The bibliometric distributions// Library Quarterly. — 1976.— Vol. 46, No. 4.— P. 416-423.
60. *Price D. J. S.* 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.
61. *Wikipedia*. —2014.— <http://en.wikipedia.org/wiki/Wikipedia>.
62. *Wikimetrics*. — 2014. —<http://www.mediawiki.org/wiki/Analytics/Wikimetrics>.
63. *Software metric*. — 2014. — [http://en.wikipedia.org/wiki/Software\\_metric](http://en.wikipedia.org/wiki/Software_metric).
64. *Flossmetrics*. — 2014.— <http://www.flossmetrics.org>.
65. *Nierop E. V.* Why do statistics journals have low impact factors?/ Statistica Neerlandica.— 2009.— Vol. 63, No. 1.— P. 52–62.
66. Not-so-deep impact// Nature.— 2005.— Vol. 435, No. 7045.— P. 1003–1004. —doi:10.1038/4351003b. PMID 15973362.
67. *Lozano G. A., Larivière V., Gingras Y.* The weakening relationship between the impact factor and papers' citations in the digital age// Journal of the American Society for Information Science and Technology.— 2012.— Vol. 63, No. 11.— P. 2140.
68. *Bergstrom C. T., West J. D., Wiseman M. A.* The eigenfactor™ metrics// Journal of Neuroscience.— 2008.— Vol. 28, No. 45.— P. 11433–11434.
69. *Bergstrom C. T.* Eigenfactor: Measuring the value and prestige of scholarly journals// College & Research Libraries News.— 2007.— Vol. 68, No. 5.
70. <http://ox.libguides.com/content.php?pid=207971&sid=1733765>.
71. [http://help.scopus.com/Content/h\\_jrnlsnip.htm](http://help.scopus.com/Content/h_jrnlsnip.htm).
72. *Kaur J., Radicchi F., Menczer F.* Universality of scholarly impact metrics// Journal of Informetrics.— 2013.
73. [http://wiki.lib.sun.ac.za/images/3/3d/2013-OA-BBLpresentation\\_METRICS.pdf](http://wiki.lib.sun.ac.za/images/3/3d/2013-OA-BBLpresentation_METRICS.pdf).
74. *J. Priem (@jasonpriem)*, I like the term #articlelevel-metrics, but it fails to imply \*diversity\* of measures. Lately, I'm liking #altmetrics., 4:28 AM - 29 Sep 2010, Tweet.
75. *Priem J., Taraborelli D., Groth P., Neylon C.* Altmetrics: A manifesto.— 2010.— <http://altmetrics.org/manifesto>.
76. A new framework for altmetrics. ImpactStory Blog. 2012-09-14.
77. *Lin J., Fenner M.* Altmetrics in evolution: Defining and redefining the ontology of articlelevel metrics// Information Standards Quarterly. — 2013.— Vol. 25, No. 2.— P. 20.

# Определение науки о данных\*

Яньюонь ЧЖУ  
(Yangyong ZHU),

*За рамками исследования правил естественного мира, отражаемого данными*

Юйнь СЯОН  
(Yun XIONG)

Школа вычислительной техники,  
Фуданский университет, г. Шанхай, Китай

Наука о данных завоевала широкое внимание в научных и промышленных кругах. Новые научные институты и организации, связанные с наукой о данных, такие как Институт наук о данных и проектирования Колумбийского университета и Центр науки о данных Нью-Йоркского университета продолжают появляться на мировой сцене. Калифорнийский университет в Беркли, Колумбийский университет, Фуданский университет и другие учредили курсы по науке о данных и программы на соискание ученой степени. Кливленд и Смит предположили, что наука о данных должна считаться независимой дисциплиной [1,2]. Компании Facebook, Google, EMC, IBM и другие ввели штатные должности для ученых, занимающихся данными. Согласно журналу *Harvard Business Review*, ученый сферы данных является представителем «самой привлекательной специальности XXI века». Сегодня имеется несколько точек зрения по вопросу определения науки о данных. Однако согласованного определения нет. Мы считаем, что научные цели новой науки, науки о данных, отличаются от других, более устоявшихся ветвей науки. Более того, научные проблемы, изучаемые наукой о данных, не исследуются естественными или общественными науками.

Наша команда работала над технологией данных и научными проектами, финансируемыми китайским правительством начиная с 1998 г., и мы применили нашу работу к наукам о жизни, здравоохранению, финансам, транспорту и другим областям (таб.). За исследуемые годы мы отметили ряд общих вопросов, связанных с данными в научных исследованиях и их применением в промышленности, особенно примечательным было сходство объектов данных. Мы осознали, что имеется значительная потребность провести отдельное исследование по самим данным, и начали изучать понятие науки о данных в 2009 г. [3]. Начиная с 2010 г. мы принимали у себя ежегодный международный симпозиум по науке о данных и dataology (iwdds.fudan.edu.cn). Симпозиум обеспечивает форум для дискуссий по вопросам науки о данных с учеными, занятыми в вычислительной технике, науках о жизни, астрономии и других областях.

За прошедшие 16 лет наше понимание науки о данных обрело более конкретную форму. Мы полагаем, что данные в киберпространстве образуют то, что мы называем природой данных [3,4]. Наука о данных является научным исследованием природы данных.

Имеется несколько современных точек зрения относительно науки о данных.

## **1-я точка зрения: Наука о данных – наука об изучении научных данных**

Комитет по научно-технологическим данным при Международном совете научных союзов (CODATA) основал журнал *Data Science Journal* (codata.org/dsj/) в 2002 г. CODATA рассматривает науку о данных как методы и технологии, используемые для проведения научных исследований посредством управления и утилизации научных данных. Так как научные данные становятся более доступными, наука о данных используется для лучшей характеристики природных интенсивно используемых современной наукой и проектированием данных. Многие дисциплины используют технологию данных, чтобы иметь дело с научными данными из их соответствующих областей. Отсюда появилась X-информатика, включающая биоинформатику, нейроинформатику и социальную информатику.

Например, ученые корпорации NuMedii, компании, занимающейся вопросами больших данных в Кремниевой долине, с помощью изучения выделения генетических данных из более 2500 примеров раковой опухоли предсказали, могут ли использоваться существующие лекарства для лечения рака яичников [5].

В качестве другого примера – математики из Гарвардского университета Эйден и Мишель изучали историю Америки, используя Ngrams в Google [6]. Они пользовались Ngrams для проведения поиска частоты использования двух фраз: «United States are» и «United States is». Результаты поиска показали, что до Гражданской войны обе фразы использовались с приблизительно одинаковой частотой, но после Гражданской войны последняя стала больше употребляемой, чем первая. Это служит показателем уровня принятия обществом США себя как единой нации до и после Гражданской войны.

\* Перевод Zhu Y., Xiong Y. Defining data science.— 2015.— <http://arxiv.org/ftp/arxiv/papers/1501/1501.05039.pdf>

### Научные проекты в областях данных, в которых принимала участие наша команда

№ п/п	Год	Название проекта	Источник финансирования
1	1998	Разработка и применение средств (инструментов) для извлечения данных	Национальная программа Китая по исследованиям высоких технологий и развитию (Программа 863)
2	2000	Разработка и применение платформы биоинформатики для целевого определения лекарств	Национальная программа Китая по исследованиям высоких технологий и развитию (Программа 863)
3	2002	Разработка и применение платформы по извлечению данных	Национальная программа Китая по исследованиям высоких технологий и развитию (Программа 863)
4	2004	Оптимизация технологии и применение сетевого сайта Shanghai World Expo	Муниципальное правительство Шанхая
5	2005	Исследование модели данных, индекса и архитектуры биологической последовательности базы данных	Национальный фонд естественных наук Китая
6	2006	Получение транскрипционного фактора запрета сайтов и композитных модулей	Национальная программа Китая по исследованиям высоких технологий и развитию (Программа 863)
7	2008	Исследование основных технологий в анализе данных для предотвращения рисков медицинского страхового финансирования	Муниципальное правительство Шанхая
8	2009	Новое поколение высокочетких изображений для интеллектуального мониторинга дорожного движения и индустриализация системы информационного обслуживания	Муниципальное правительство Шанхая
9	2010	Исследование алгоритма извлечения на основе характеристик для построения биологических сетей	Национальный фонд естественных наук Китая
10	2011	Разработка и операции облачной платформы для фондовой бирже	Национальная программа поддержки естественных наук и технологий
11	2012	Исследование алгоритма извлечения для особых групп	Национальный фонд естественных наук Китая
12	2012	Исследование медицинских больших данных	Муниципальное правительство Шанхая
13	2012	Теоретическое исследование больших данных по городскому дорожному движению	Муниципальное правительство Шанхая
14	2013	Исследование теорий науки о данных	Муниципальное правительство Шанхая
15	2014	Исследование технологий больших данных для шанхайской кораблестроительной отрасли	Муниципальное правительство Шанхая

С этой точки зрения, данные главным образом относятся к данным, порождаемым и используемым в научных исследованиях. Это подчеркивает, что наука о данных является управлением, обработкой и использованием научных данных в целях поддержки научных исследований, т.е. современное широко известное, интенсивно использующее данные научное исследование или четвертая парадигма научного исследования [7].

#### 2-я точка зрения: Наука о данных – наука изучения деловых (бизнес) данных.

В 2010 г. Лукидес обсуждал, что такое наука о данных, утверждая, что наука о данных должна скорее способствовать созданию продуктов данных, а не являться работой по простому применению данных [8]. В 2013 г. Провост и др. отмечали, что «извлечение знания из данных для решения проблем бизнеса» является одним из фундаментальных понятий науки о данных [9].

В условиях обеспечения поддержки методологии ВІ исследование составляет значительную часть работы, осуществляемой многими учеными, связанными с данными. Для усиления эффекта скажем, что большая доля практиков сферы ВІ перешла в ученые, занимающиеся данными. Amazon, Google, LinkedIn, Facebook и другие интернет-компании ввели должности для занимающихся данными ученых и создали команды науки о данных. Эти специалисты по данным изучают и анализируют бизнес-данные, чтобы обеспечивать службы управления принятием решения. Например, Amazon использует совместный фильтр для генерирования высококачественных рекомендаций относительно продукта, а Facebook использует характеристику «Люди, с которыми вы можете быть знакомы» для рекомендации дружеских связей.

С этой точки зрения, получение знания от бизнес-данных, чтобы принять решение, является одним из аспектов науки о данных. Это подобно тому, над чем работают ученые сферы ВІ. По этой причине многие

ученые ВІ также называются учеными, изучающими данные. Тем не менее, по сравнению с проблемами ВІ наука о данных фокусируется на общих вопросах в анализе различных бизнес-данных, т.е. вопросах методологии по ВІ.

### **3-я точка зрения: Наука о данных – интеграция статистики, компьютерной технологии и искусственного интеллекта (ИИ)**

Эта точка зрения часто возникает в дискуссиях по поводу того, кем являются ученые, занимающиеся данными. Вообще считается, что такие ученые должны иметь навыки в области статистики, вычислительной техники, ИИ и связанных с ними областей, и что они не являются отдельными людьми, специализирующимися в одной области, поскольку команды состоят из статистиков, компьютерных специалистов, экспертов в области ИИ и экспертов других релевантных областей.

Например, команды ученых, специализирующихся на данных, в Google и Facebook включают статистиков, компьютерных специалистов, специалистов в области ИИ и экспертов в других релевантных областях.

Эта точка зрения проста: так как статистика, компьютерная технология и ИИ – используются для обработки и анализа данных, то все они являются неотъемлемой частью науки о данных.

### **4-я точка зрения: Цель науки о данных – решать научные и деловые проблемы путем извлечения знания из данных**

В 2013 г. Дхар обсуждал применение науки о данных с точки зрения бизнеса и научных исследований [10]. Он определил науку о данных как «изучение обобщенного извлечения знания из данных» [10]. Он также указал на то, что ученый сферы данных должен обладать исчерпывающими навыками, охватывающими статистику, машинное обучение, ИИ и управление базами данных, и глубоким пониманием разработки проблемы. Эта точка зрения может рассматриваться как объединение первых трех.

#### **Чем является наука о данных?**

Основные идеи, лежащие в основе вышеуказанных определений, состоят в том, что наука о данных используется для получения знания из данных в некоторых

релевантных областях и обеспечения поддержки существующей схемы научного исследования и управления принятым решением. Однако вся описанная выше работа пока является недостаточной для утверждения науки о данных в качестве новой, уникальной ветви науки. Это происходит потому, что объектами изучения являются предметы естественного мира, и их научные проблемы также изучаются существующими научными областями.

С развитием цифрового оборудования предметы естественного мира все больше хранятся в киберпространстве в форме данных. Данные вводятся, порождаются и создаются в киберпространстве множеством способов и становятся все более и более разнообразными, сложными и выходящими из-под контроля человека. Все больше данных являются неизвестными или плохо понимаемыми людьми. Данные в киберпространстве уже показывают черты независимого мира, подобного естественному миру, поэтому все данные в киберпространстве здесь относят к природе данных.

Следует отметить, что в киберпространстве имеется два типа данных. Первый – это данные, представляющие вещи в естественном мире, называемые здесь реальными данными. Примером служит личная информация, которая является данными, представляющими личные характеристики. Вторым типом являются данные, не представляющие вещи в естественном мире, так называемые виртуальные данные. Виртуальные данные означают, что примеры таких данных не имеют связи с естественным миром. Примерами этого являются компьютерные вирусы, которые не являются ни вирусами в естественном мире, ни представлением данных реальных вирусов; вместо этого они только существуют в киберпространстве (рис. 1).

Формация природные данные производит новые объекты исследования и ставит новые научные проблемы. Эти новые объекты исследования не являются вещами, существующими в естественном мире или сообществе людей, а скорее существующими в природных данных, т.е. данными. Имеются новые научные вопросы о природных данных. Каков размер природных данных? Каков темп роста глобальных данных? Как распространяются данные в киберпространстве? Как следует определять достоверность природных данных? Ни один из этих вопросов не рассматривается естественными или общественными науками. Эти новые научные вопросы следует изучать с помощью новой науки.

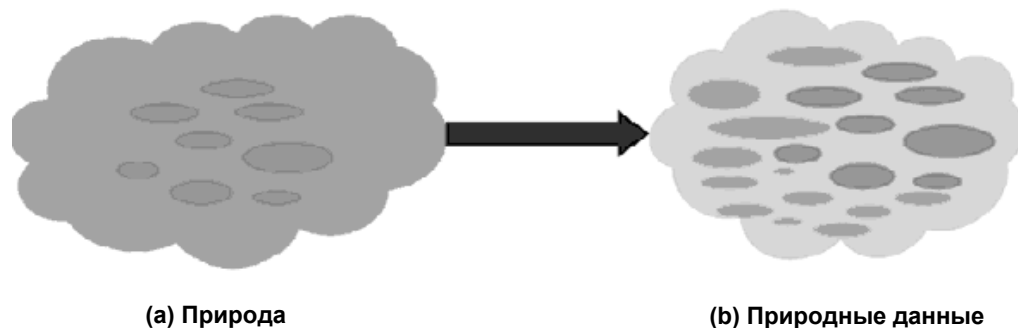


Рис. 1. Реальные данные и виртуальные данные

*Примечание:* эллипсы в правой части рисунка (b) означают реальные данные, представляющие вещи в естественном мире. Среди них более темные эллипсы соответствуют тем вещам в естественном мире, которые хранятся в киберпространстве в форме данных (т.е. эллипсам в левой части рисунка (a)): более светлые эллипсы правой части рисунка (b) соответствуют тем вещам в естественном мире, которые должны храниться в киберпространстве (т.е. в «облаке» левой части рисунка (a)). Совокупность более светлых эллипсов правой части рисунка (b) означает виртуальные данные, не представляющие вещи в естественном мире, т.е. примеры такого рода данных, которые не имеют связей с естественным миром.



Рис. 2. Темы исследования науки о данных

Наука о данных является наукой изучения природных данных и самой наукой данных. На основном уровне она включает извлечение знания из данных. Так как некоторые данные в природных данных представляют реальные вещи, знание, получаемое из этих данных, можно использовать в естественных и общественных науках. Этот тип работы считается наукой о данных согласно точкам зрения 1- 4. Однако это только одна часть исследования науки о данных.

Наука о данных здесь определяется следующим образом:

Наука о данных является теорией, методом и технологией изучения природных данных. Она имеет два основных компонента.

**Первым компонентом является изучение моделей и правил самих данных. Его цель - изучить природные данные и научные проблемы, связанные с природными данными.** Оно не учитывает значение данных в естественном мире.

**Вторым компонентом служит изучение правил естественного мира, отражаемого данными, т.е. изучение естественного мира, осуществляемое через изучение данных.** Например, цель выполнения исследования данных, представляющих поведение человека, – изучить поведение этого человека.

Как упоминалось выше, изучение данных существует уже некоторое время, и методы данных, такие как извлечение данных, развиваются. Тем не менее, научное сообщество в сфере науки о данных нуждается в установлении фундаментальной теории и основных методов научного наблюдения, измерения и дальнейшей разработки технологий данных. На рис. 2 показаны основные темы науки о данных.

### Заключение

Наука о данных завоевывает все более широкое внимание, но нет согласованной точки зрения относительно того, что такое наука о данных. Объекты исследования и научные проблемы этой новой науки не должны перекрываться устоявшимися науками. В данной статье наука о данных определяется как наука изучения природных данных. Мы считаем, что это самое логичное и точное определение науки о данных, и оно включает основные части определений точек зрения 1- 4.

### Литература

1. *Cleveland W. S.* Data Science: An action plan for expanding the technical areas of the field of statistics// *International Statistical Review.* — 2001.— Vol. 69, No 1.— P. 21-26.



2. *Smith F. J.* Data Science as an academic discipline// Data Science Journal.—2006. — Vol. 5.— P. 163-164.
3. *Zhu Y. Y., et al.* Data explosion, data nature and dataology// Proceedings of International Conference on Brain Informatics. — 2009.
4. *Zhu Y. Y., Xiong Y.* Dataology and data science: Up to now. — 2011.— [http://www.paper.edu.cn/en\\_releasepaper/content/4432156](http://www.paper.edu.cn/en_releasepaper/content/4432156).
5. *May M.* Life science technologies: Big biological impacts from big data// Science (Science Products, Product Articles).— 13 June 2014.— [http://www.sciencemag.org/site/products/lst\\_20140613.xhtml](http://www.sciencemag.org/site/products/lst_20140613.xhtml).
6. *Aiden E., Michel J. B.* Uncharted: Big data as a lens on human culture.— Penguin Group, 2013.
7. *Hey T., et al.* The fourth paradigm: Data intensive scientific discovery// Microsoft Research. — Oct. 2009. — <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>.
8. *Loukides M.* What is data science? An O'Reilly Radar Report. — 2010. — <http://radar.oreilly.com/2010/06/what-is-data-science.html>.
9. *Provost F., Fawcett T.* Data science and its relationship to big data and data-driven decision making// Big Data.— 2013.— Vol. 1, No. 1.— P. 51-59.— <http://online.liebertpub.com/doi/pdfplus/10.1089/big.2013.1508>.
10. *Dhar V.* Data Science and prediction// Commun. ACM. — 2013. — Vol. 56, No.12 (Dec. 2013).— P. 64-73.

# Социальные, политические и правовые аспекты извлечения текстов и данных\*

**Мишель БРУК**  
(Michelle BROOK)

Организация «The Content Mine»,  
Великобритания

**Питер МЮРРЕЙ-РАСТ**  
(Peter MURRAY-RUST)

Кембриджский университет,  
Великобритания

**Чарлз ОППЕНХАЙМ**  
(Charles OPPENHEIM)

Нортхемптонский университет  
и Университет Роберта Гордона,  
Великобритания

*Идеи извлечения текстов или данных (TDM - textual or data mining) и последующего анализа уходят корнями на сотни, если не тысячи, лет назад. Первоначально проводимый вручную анализ текстов и данных долгое время являлся средством, дающим возможность получить новые знания, взятые из совокупностей текста. Однако для открытия потенциальных выгод TDM следует преодолеть ряд не относящихся к технологии барьеров. Они включают правовую неопределенность, являющуюся результатом сложного авторского права, прав баз данных и лицензирования, факт, что некоторые издатели не всегда учитывают возможности, которые TDM предлагает научному сообществу, и отсутствие осведомленности относительно TDM среди многих ученых, а также пробел в приобретении навыков.*

## Введение

Идеи анализа текстов или данных возникли столетия, если не тысячелетия, назад. Вначале анализ текстов и данных производился вручную и долгое время являлся средством, позволяющим извлекать новые знания из совокупностей текстов и данных.

Развитие компьютеров дало возможность автоматизировать вовлеченные процессы, включая информационный поиск, исследование индексирования, чтение и анализ существующей в цифровом формате литературы. Эти автоматизированные процессы, часто рассматриваемые как **извлечение текста и данных**, имеют много ежедневных применений, включая лежащую в основе поисковых устройств Интернета технологию и целевое рекламирование, разнообразие способов получения доступа к библиотечным фондам и пожелания. Существует много потенциальных применений TDM в рамках научного учреждения, включая биомедицинское, биологическое и химическое исследование, где информация на уровне генов (общая информация), названия видов, химические соединения и статистические значения также могут быть извлечены. Кларк [1] предоставляет полезное введение к данной теме.

Хотя основанные на компьютере методы возвращают нас к началу 1980-х гг., неудивительно, что развитие TDM в качестве области очень во многом серьезно по-

лагается на усовершенствования в сфере вычислительной аппаратуры, включая стоимость запоминающих устройств с небольшим объемом памяти и стоимость карточных перфораторов (CPU), развитие высокоскоростных сетей, а также инноваций в программном обеспечении, в том числе средства обработки естественного языка и оптическое распознавание знаков.

Любой подвергающийся анализу контент должен существовать в том, что называется «машиночитаемым форматом», в одном из нескольких форматов, известных как XML или JSON, которые позволяют компьютеру читать и понимать содержащуюся информацию. PDF, часто смешанный язык (lingua franca) научных журналов, будучи способным представлять стандартизированные форматы для чтения человеком, не являются машиночитаемыми, а также пригодными для целей TDM, эта работа должна передаваться в другой цифровой форме [2]. Эта форма часто привычна и специфична для задаваемого научного вопроса и является наиболее подходящим средством для ответа на него.

Неудивительно, что существует обширная литература, описывающая исследование и применения TDM, а также технические проблемы. Однако на сегодняшний день сравнительно мало написано по поводу имеющихся политических, социальных и правовых барьеров. Первая правовая проблема, связанная с TDM, относится к авторскому праву, правам в отношении баз данных и лицензирования\*, для которых имеются политические

\* Перевод Brook M., Murray-Rust P., Oppenheim Ch. The social, political and legal aspects of text and data mining (TDM). — <http://www.dlib.org/dlib/november14/brook/11brook.print.html>

\* См. Creative Commons, about the licences

решения. Отсутствие осведомленности и соответствующий технологический пробел между многими средствами TDM и навыками ученых также действуют в качестве барьера.

Несмотря на существование технологических и правовых барьеров, имелся рост интереса к TDM в последние годы; сюда входит создание первого, финансируемого обществом центра по извлечению текста, Национального центра по извлечению текста в Манчестере в 2004 г.; TDM является предметом многочисленных отчетов, включая отчет JISC 2012 г. [3] и развитие **Google Ngram**. Такой рост интереса – не просто отражение совершенствований в методах и технологиях. Есть осознание того, что технологии TDM обеспечивают широкие экономические и общественные возможности, включая «растущую эффективность ученого; открытие скрытой информации и развитие нового знания; исследование новых горизонтов; усовершенствование научной и доказательной базы; улучшение научного процесса и качества» [3].

Однако, чтобы раскрыть потенциальные преимущества TDM, необходимо преодолеть ряд нетехнологических барьеров. Они включают правовую неопределенность, являющуюся результатом сложного авторского права, прав баз данных и лицензирования; тот факт, что некоторые издатели не всегда принимают возможности, которые TDM предлагает академическому сообществу, и отсутствие осведомленности и технологических навыков среди многих ученых в отношении TDM. Это не единственные барьеры в принятии TDM; однако такие темы, как вопросы и проблемы сохранения тайны медицинских данных; тот факт, что качество данных может быть плохим, например, данные, содержащиеся в различных системах, отсутствующие данные, недостоверные данные и не стандартизированные данные; то, что иногда возникают модели, которые на самом деле относятся к выбранным наугад колебаниям; и тот факт, что создание сложного устройства TDM требует существенных инвестиций в системы информационной технологии, создание банков данных и укрепление экспертизы – в данной статье не рассматриваются.

### **Авторское право, право баз данных, лицензии и TDM**

Основными правовыми проблемами, стоящими перед любым, кто хочет осуществить TDM, являются закон об авторском праве, закон о базах данных и закон о контракте (лицензирование работы)\*\*. В то время как рассматриваемые проблемы релевантны во всем мире, комментарии, приведенные ниже и касающиеся правового положения, специфичны для законодательства Великобритании, и читатели не должны полагать, что они применимы в других юрисдикциях. Однако в этих юрисдикциях будут схожие моменты.

Закон *об авторском праве*, как многие утверждают, гласит, что каждому, создающему и записывающему в любом виде новую творческую работу (т.е. не скопированную где-нибудь в другом месте), автоматически гарантируется защита авторского права на эту работу. Такая защита обеспечивает владельцу право давать или не давать разрешение на совершение определенных «ограниченных действий» любой третьей стороной; эти действия включают копирование, адаптацию работы, по-

вторное распространение (например, путем ее публикации в сети) всей или «значительной части» оригинальной, защищенной авторским правом работы, а также перевод работы на другие языки.

На практике закон об авторском праве гораздо сложнее, чем говорится выше. Простые факты (которые включают многие данные), а также короткие предложения не являются предметом авторского права. То, что классифицируется как факт или короткое предложение, остается неясным и противоречивым и часто устанавливается при помощи судебного рассмотрения. Недавний прецедент Европейского суда\*\*\* подтвердил, что предложение из 11 слов допускает, чтобы быть защищенным авторским правом, хотя это не подразумевает, что все подобные предложения будут защищены авторским правом.

Поскольку копирование считается нарушением права интеллектуальной собственности владельцев авторского права, то это ведет к некоторой правовой неясности в отношении процесса TDM. Некоторые процессы TDM, из PDF, могут требовать создания временного промежуточного файла. Являются ли такие файлы «правовыми» (законными) – вопрос дебатов между теми, кто поддерживает, и теми, кто противостоит широкому использованию TDM. Является ли промежуточный файл эквивалентом временного файла из Adobe Acrobat Reader? Или же это производная работа, а следовательно, и нарушение авторского права?

Эти вопросы важны. Поскольку большая часть научной литературы в настоящее время находится под защитой авторского права, без отказа прав на лицензию, мы можем предположить, что большинство литературы, которая будет использоваться для TDM, должно находиться под защитой авторского права. В то время как существуют некоторые так называемые исключения к авторскому праву\*\*\*\*, которые позволяют третьим сторонам копировать или осуществлять ограниченные действия без оплаты или без получения разрешения, и хотя такие исключения в некоторых странах, таких как США, являются достаточными, чтобы позволить извлечение контента, ученые могут быть вынуждены подвергнуть проверке закон, который недостаточно хорошо определен для таких видов деятельности. В Великобритании до недавнего введения изменения к закону об авторском праве имели место многочисленные исключения к этому закону, однако их формулировка была либо неясной, либо не бесспорно применялась к TDM.

Результатом этого явился правовой барьер, который может препятствовать нерасположенным к риску ученым (или их организациям) полагаться на исключения к авторскому праву в осуществлении их исследования, и вместо этого им надо обращаться к издателям (т.к. большинство работ, которые они хотят использовать, действительно принадлежат издателям) за разрешением на TDM для определенного проекта, для специфического предварительно очерченного вопроса. Иногда такое разрешение будет охотно предоставляться бесплатно, но многие издатели хотят необоснованно высокой платы и/или ставят ограничения на то, что можно делать с их материалами, и/или устанавливают чрезвычайно длительное время для принятия решения относительно того, как ответить на запрос; таким образом,

\*\*\* См. Case C-5/08, Infopaq Int'l A/S v. Danske Dagblades Forening [2009] ECR I-6569

\*\*\*\* Ими часто являются так называемые исключения «честного использования» или «честного ведения дел» в рамках Великобритании.

\*\* Дополнительно существует ряд связанных проблем, таких как конфиденциальность и защита данных, которые мы здесь не исследуем.

практически получить разрешение почти невозможно. Имеется ряд примечательных исключений, таких как издательства «Springer» и «The Royal Society», которые приняли на себя благородный подход в отношении предоставления разрешений ученым осуществлять TDM, но увы, они все еще в меньшинстве. В любом случае ученые, осуществляющие TDM, должны обращаться к многочисленным издателям, каждый из которых имеет свои отношения, условия и время ответа на такие запросы. Для ученого это очень затратно и будет иметь значительное влияние на принятие TDM в качестве практики или (как показывает история) тормозить ученых в процессе совместного использования результатов их исследования с помощью TDM, таким образом ограничивая потенциальные выгоды, которые обеспечивает данная технология.

### Недавние изменения в законодательстве Великобритании

В 2010 г. премьер-министр Великобритании подготовил доклад, нацеленный на то, чтобы рассмотреть вопрос о том, готова ли система интеллектуальной собственности Великобритании к задаче поддержки инновации и роста (the Hargreaves report) [4]. Одна из основных рекомендаций (Гл. 5,26) состояла в том, что правительство Великобритании должно ввести исключение к авторскому праву, разрешающее использовать аналитиков в некоммерческих целях. Не ясно, где Харгривз почерпнул эту идею, но она частично должна основываться на свидетельстве, представленном библиотечным и научным сообществами Великобритании [5]. В декабре 2012 г. Бюро интеллектуальной собственности Великобритании (IPO) провело оценку влияния введения такого «исключения для копирования (воспроизведения) работ в целях использования аналитиками текста и данных». Для этого оно определило «текст и данные, а также методы аналитических данных» как методы для: «извлечения данных из существующей электронной информации, установления новых фактов и взаимосвязей, создания новых научных результатов из предшествующего исследования. Эти новые методы включают копирование предшествующих работ как часть процесса извлечения данных» [6]. Оно (Бюро) не пыталось применить картину подобного экономического влияния этого изменения к законодательству, но отмечало, что в дальнейших условиях изменение приведет к «социальной инновации и дальнейшему простору для основных экономических выпрышек».

1 июня 2014 г. Парламент Великобритании принял законодательство по изменению закона относительно TDM, в котором говорится:

«Снятие копии работы лицом, имеющим к ней законный доступ, не нарушает авторское право в следующих случаях:

а) копия делается, чтобы лицо, имеющее законный доступ к работе, могло осуществить вычислительный анализ чего-либо, приведенного в работе, с единственной целью – проведение исследования для некоммерческой цели;

б) копия сопровождается соответствующим уведомлением (до тех пор пока это не будет возможно по практическим или другим причинам)».

Критическое дополнение констатирует:

«До тех пор пока условие контракта подразумевает предотвращение или ограничение снятия копии, которая, в силу этого раздела, не будет нарушать авторское право, это условие не может проводиться в жизнь».

Важность этого изменения к закону нельзя слишком переоценивать. Во-первых, одним росчерком пера оно

прекращает споры в отношении того, может или не может TDM быть разрешено при предшествующих исключениях к авторскому праву Великобритании. Во-вторых, оно делает попытки с помощью издателей навязать их собственные условия ученым Великобритании, что вызывает трату времени и денег каждого. В-третьих, оно предлагает ролевою модель для других стран в целях формулировки законодательства, чтобы помочь продвинуть эти полезные методы. Наконец, по крайней мере временно, оно предлагает ученым других стран, проводящим TDM, способ избежать беспокойства относительно ограничений и нарушения авторского права с помощью совместной работы с учеными Великобритании, которые имеют законный доступ к релевантным необработанным материалам, позволяя ученым, осуществляющим TDM, проводить работу, зная, что она не включает правового риска.

Ясно, что многочисленные ограничения для ученых Великобритании с помощью одного усилия были отодвинуты. Нельзя сказать, что положение является совершенным. Все еще существует неразбериха относительно того, как широко может интерпретироваться термин «некоммерческое исследование». Понятно, что ученый, нанятый на службу фармацевтической компанией, не может претендовать на то, что выполняет «некоммерческое» исследование, но как быть с ученым в академической сфере, который получил финансирование от фармацевтической компании или был взят коммерческой компанией для проведения исследования с помощью TDM? Или независимый ученый, планирующий опубликовать их работу в блоге, который получает доход от помещения рекламы? На эти вопросы можно ответить только с помощью судебных разбирательств – иногда таких, в которых вынуждены участвовать владельцы прав и ученые, осуществляющие TDM, и до тех пор пока эти вопросы не разрешены, имеется много ученых, которые не решаются в полной мере использовать указанные средства.

Права баз данных схожи с авторским правом, но более ограничены, и применяются к определенным коллекциям данных (данных, включающих множество сведений, таких как слова, цифры или изображения, но не фильмы или звуковые записи). Защита прав баз данных является сложной, и здесь приводится только упрощенное изложение. База данных может обладать авторским правом (и быть подверженной упомянутым выше пунктам относительно закона об авторском праве), если массив и организация материалов в базе данных включает интеллектуальное усилие. Другими словами, массив данных может быть защищен авторским правом, если имеется достаточно креативности, вовлеченной в презентацию или организацию данного набора, хотя пункты основных индивидуальных данных остаются свободными от любой защиты.

Таким образом, полностью исчерпывающая база данных документов, где не осуществляется отбор, не может обладать авторским правом. Но если здесь имеет место интеллектуальное усилие в проведении отбора, тогда авторское право применяется. Кроме того, и что выделяется совершенно особо, база данных может обладать правом баз данных, если тратится усилие любого рода при получении, подтверждении и представлении данных.

Следовательно, база данных может обладать авторским правом, правом базы данных, или может не иметь вообще никаких связанных с этим прав в первую очередь в зависимости от того, с какой целью она была создана. Что касается авторского права, то здесь закон часто является непонятным и действует в качестве расхолаживающего эффекта и барьера для многих лиц.

*Лицензии/контракты* представляют собой важные черты правового ландшафта TDM. Владельцы авторского права и/или прав баз данных часто будут должны давать разрешение третьим сторонам на чтение, загрузку, повторное распространение или, иначе, на использование своих материалов, как правило, за плату (что может быть важным в случае больших научных издательств, владеющих широким диапазоном ресурсов). Такие лицензии будут неизбежно навязывать определенные соглашения и условия – ограничения – на то, что лицензент может делать с лицензированным материалом. Все в возрастающей степени мы наблюдаем понимание в Creative Commons licences, которое дает авторам возможность разрешать третьим сторонам использовать их работу в заранее определенных действиях. Однако даже эти сравнительно либеральные лицензии навязывают некоторые соглашения и условия пользователю, такие как необходимость подтвердить источник\*\*\*\*. Многие, не относящиеся к Creative Commons, лицензии не позволяют проводить TDM (хотя эти лицензии, конечно, не принимаются во внимание исключениями к авторскому праву в рамках определенной страны, такой как Великобритания или Япония, которая имеет схожие исключения с Великобританией в отношении TDM).

В то время как некоторые государственные финансирующие организации, такие как советы по научным исследованиям Великобритании, в настоящее время требуют использования лицензий Creative Commons (CC) для научной продукции, произведенной с помощью их финансовых средств, многие финансирующие научные исследования учреждения по всему миру пока еще этого не делают. Использование этих лицензий может быть очень полезным для ученых, желающих участвовать в применении TDM, в тех странах, где нет соответствующих исключений к авторскому праву. Машиночитаемый формат лицензий является решающим для TDM, хотя многие издатели все еще в полной мере используют возможности, которые предлагают машиночитаемые лицензии CC. Более того, распространению лицензий следует сопротивляться, особенно в тех случаях, где новые представленные лицензии не взаимодействуют с теми, которые уже существуют. Это особенно важно в международном контексте, принимая во внимание неясности в интерпретации закона об авторском праве в рамках различных юрисдикций. При наличии лицензий Creative Commons, фактического стандарта в лицензировании Открытого доступа и в растущей степени в рамках открытых данных, возникает вопрос, почему издатели испытывают необходимость в попытке навязать контроль в этом стиле работы и почему многие из этих лицензий оформляются как «Открытый доступ», несмотря на признание определения открытого доступа BOAI (BOAI definition of Open Access).

Свежий пример распространения лицензий можно наблюдать на действиях Международной ассоциации научных, технических и медицинских издателей, которая недавно распространила свой собственный набор лицензий [7]. Во введении к этим лицензиям говорится, без какого-либо поддерживающего свидетельства, что «сложность интересов и обязательств владельцев капитала означает, что ни одно навязанное извне решение не может быть эффективным, будь оно национальным исключением или чем-то еще: следовательно, решения должны выработываться с помощью диалогов многих

владельцев капитала, которые уважают интересы владельцев капитала, проводятся в прозрачной манере и нацелены на модели, являющиеся простыми и беспристрастными». Такие выработанные издателем лицензии на проведение TDM являются предметом влияния сильной стороны, а также оправданного критицизма со стороны библиотечных [8] и научных сообществ, поскольку предлагаемые лицензии представляются слишком ограничивающими и навязывают пристрастные (а в случае Великобритании не проводимые в жизнь) давления на свободу ученых применять TDM.

### **Что могут сделать политики и разработчики политики**

Во многих странах закон, связанный с TDM, является неясным, и это будет сдерживать ученых от его принятия. Авторское право и права баз данных должны развиваться, чтобы отражать современные практики и технологии, а не сдерживать развитие нового научного исследования. Там, где имеются неясности, они должны быть убраны. Например, при современном законодательстве Европейского Союза (ЕС) страны могут ввести исключения для некоммерческого использования TDM при проведении исследования, но на сегодняшний день только Великобритания использует это преимущество. Сегодня нет никаких изменений на уровне ЕС, хотя они считают ЕС широким полем для исключений относительно TDM, Республика Ирландия также предусматривает такое изменение в своем законодательстве. За пределами Европы авторы данной статьи слышали только об одной или двух странах Дальнего Востока, которые ввели такие исключения. Это основной барьер для применения TDM по всему миру.

Ясно, что Великобритания является ведущей в этой области, тем не менее, остается неясность относительно того, что является «коммерческим исследованием», а что – нет, может потребоваться несколько судебных дел, чтобы прояснить ситуацию, хотя неопределенность, вероятно, имеет расхолаживающий эффект до тех пор, пока не будут разрешены эти вопросы. Финансирующие научное исследование организации могут поддержать использование лицензий Creative Commons для выхода научной продукции, являющейся результатом их финансирования, они (финансирующие организации) придерживаются правильной позиции в работе со своими носителями капитала по вопросу проблем, сталкивающимся с распространением лицензий.

### **Издатели не используют возможности TDM**

Понятно, что некоторые издатели сопротивляются тому, чтобы отказаться от контроля, который они в настоящее время, как правило, осуществляют над тем, как проводится TDM относительно материалов, авторским правом на которые они владеют. Приводится типичное заявление Ричарда Моллета, руководителя Ассоциации издателей Великобритании: «Мы поддерживаем извлечение контента. Это может хорошо работать только в том случае, если мы вовлечены в процесс и управляем доступом»\*\*\*\*\*. Руководство к законодательству, введенному в ответ на законодательство Харгривза, особо констатирует, что издатели могут применять только «технологические меры в отношении сетевых работ, которые необходимы для поддержания безопасности или стабильности, но не способны усилить условия контрактов,

\*\*\*\* См. Creative Commons, about the licences.

\*\*\*\*\* См. R. Mollet. Statement to Houses of Parliament

которые стремятся не допустить или беспричинно ограничить извлечение текста или данных», будет интересно посмотреть, признают ли издатели, например, что требование ученых, применяющих TDM, использовать только «одобренные» API, не является законным.

### Как издатели могут помочь ученым в осуществлении TDM?

Издатели могут сделать многое, чтобы помочь ученым использовать TDM. Принимая во внимание недавние изменения к закону Великобритании и тот факт, что другие страны теперь либо имеют схожие законы или предполагают их ввести (и действительно существует давление со стороны директивы ЕС, предлагающей ввести такое исключение в рамках ЕС повсеместно), издатели могут помочь осуществлению исследования с помощью TDM путем прекращения усилий по «контролированию» процесса, а скорее всего с помощью прагматического адаптирования следующих политик:

1. Предложить ученым всего мира такую же свободу, какая сейчас обеспечена ученым Великобритании при проведении TDM в целях некоммерческого использования пока пользователь имеет законный доступ к оригинальным материалам, т.е. следовать примеру издательства «Springer» и «The Royal Society»\*\*\*\*\*.

2. Зарабатывать расположение среди сообщества ученых, осуществляющих TDM, путем предложения дружественных по отношению к пользователю API (конечно, без ТРЕБОВАНИЯ того, чтобы ученый их использовал), бесплатного совета и дискуссионных форумов для обмена опытом и идеями в теории и практике TDM – иначе говоря, путем адаптации политики радужного одобрения и поощрения изменения, а не пытаться сопротивляться этому.

3. Разрабатывать четкие согласованные положения относительно того, какие типы научного исследования считать «некоммерческими», а какие – «коммерческими». Это может быть успешно сделано объединенной рабочей группой издателей, исследователей, применяющих TDM, и JISC. Существует прецедент такого типа сотрудничества – **Решения издателей и библиотек** (PALS – Publisher and Library Solutions), разработавшего множество полезных согласованных заявлений за многие годы.

Не секрет, что коммерческие научные издатели в настоящее время не имеют репутации среди ученых. Это вызвано циничными политиками относительно открытого доступа (издатели используют этот термин, когда то, что предлагается, вообще не является открытым доступом) и так называемым «двойным взиманием» за «золотые» и на основе подписки журналы. Недавнее изменение в законе Великобритании предоставляет издателям «золотую» возможность улучшить свою репутацию. Но поймут ли они это?

### Осведомленность среди ученых и технологический пробел

Авторам кажется, что наблюдается отсутствие осведомленности относительно технологий TDM среди многих ученых – что возможно сделать при наличии этих технологий и как использовать эти технологии. Современные ученые, осуществляющие TDM, очень сведущи в плане технологии, но еще предстоит проделать работу по развитию существующих средств с целью сделать их более легкими для использования теми, кто имеет

меньшее знание. Пока Content Mine начинает приспосабливать что-то из имеющегося в наличии, развивая обучающие рабочие семинары по TDM, другие организации, такие как Software Carpentry, проводят семинары, чтобы помочь ученым стать более уверенными в техническом плане, но многое предстоит еще сделать. Что-то должно прийти из широкого сообщества TDM, которое должно помочь ликвидировать пробелы в знании, умении и осведомленности, однако финансирующие стороны и учреждения также несут ответственность за то, чтобы гарантированно помочь в обучении исследователей соответствующим навыкам и технологиям; это достигается путем проведения семинаров, объединяющих обучающие сессии с сессиями по докторским квалификациям, и путем выпуска информации для чтения и понимания учеными.

### Заключение

Основные барьеры в развитии понимания TDM не являются техническими, в первую очередь – это отсутствие осведомленности и навыков среди ученых. Они относятся к правовым вопросам в сфере авторского права и прав баз данных, а также к некоторым выборам политики ограничения, применяемой издателями по вопросу доступа, например, к API. Все эти проблемы решаемы, но требуют не технических решений.

В Великобритании имеются новые исключения к авторскому праву относительно TDM. Другие страны должны позаботиться о том, чтобы ввести подобные исключения, возможно, используя формулировку из законодательства Великобритании в качестве шаблона, разъясняющую, что TDM разрешено в рамках их юрисдикции. Издатели должны работать с научным сообществом TDM в целях развития согласованных заявлений по вопросу, какие типы научного исследования они признают «некоммерческими», а какие – «коммерческими», и предотвратить любой возможный «охлаждающий эффект», вызванный неясностью относительно этих терминов. Финансирующие организации и учреждения должны изучать, как обучать технике TDM заинтересованных ученых, включая возможности обучать этому в рамках программ на получение докторской степени. В Великобритании мы пошли так далеко, насколько нам позволяет политика. Что нам сейчас необходимо сделать, так это показать важность TDM, поддержать другие страны в стремлении разрабатывать такие же исключения к авторскому праву и показать ученым, что TDM стоит того, чтобы потратить их время.

### Литература

1. Clark J. Text mining and scholarly publishing.— 2012.
2. Hobbs J.R., Walker D.E., Amsler R.A. Natural language access to structured text/Proceedings of COLING' 82, the 9 th Conference on Computational Linguistics. —1982.— Vol.1.— P. 127-132.
3. McDonald D., Kelly U. The value and benefits of text mining.— 2012.
4. Hargreaves I. Digital opportunity.— 2011.
5. LACA Response to Independent Review on Intellectual Property and Growth.— 2011.
6. UK Intellectual Property Office. Economic impact of recommendations.— 2012.
7. STM. Text and data mining: STM statement and sample licence.— 2014.
8. Liber News. — 2014.

\*\*\*\*\* См. подробности политик относительно TDM, осуществляемых издательствами «Springer» и «The Royal Society».

# Ответ на поиск: предложение использовать поисковую информацию для лучшего представления контента сетевых сайтов\*

**С. Равиндранат ШОУДАРИ**  
(C. Ravindranath CHOWDARY),

**Анил Кумар СИНГХ**  
(Anil Kumar SINGH),

**Анил НЕЛАКАНТИ**  
(Anil NELAKANTI)

Индийский технологический институт,  
Бенаресский индийский университет,  
г. Бенарес, Индия

*Предполагается, что поиск и управление контентом должны быть взаимоисключающими. В данной статье говорится, что они не должны быть такими. В обычном сценарии информационного поиска некоторая информация о запросах, ведущая к сетевому сайту (благодаря «откликам» или «посещениям»), доступна администратору сервера соответствующего сетевого сайта. Эта информация может использоваться для лучшего представления контента на сетевом сайте. Более того, предполагается, что больше информации может быть разделено между информационной системой и провайдером контента. Это будет стимулировать контент-провайдера (любого сетевого сайта) иметь более динамичную презентацию контента своего сайта, т.е. в соответствии с направлениями запросов и без нарушения конфиденциальности запрашивающего пользователя. Результатом будет лучшая синхронизация между поисковыми системами и контент-провайдерами в целях улучшения пользовательского опыта сетевого поиска. Это также даст возможность контент-провайдеру заявить о себе в этом процессе, учитывая, что контент-провайдером является тот, кто знает гораздо больше о контенте, чем поисковая система. Это к тому же подразумевает, что представление контента может меняться в ответ на запрос. В конце концов, пользователь будет иметь возможность находить релевантный контент более легко и быстро.*

## Введение

Системы информационного поиска (ИП) становятся существенной частью повседневной деятельности миллионов людей, и с годами они сохраняют свое превосходство. Одной из причин такой важности надежной системы ИП является количество данных, доступных в сети, и темп роста. Число сетевых сайтов соответствующим образом увеличилось с одного в 1991 г. до более одного миллиарда в сентябре 2014 г.\*\* Одновременно растет число пользователей, помогающих ведущим национальным информационным службам. Рост сетевого использования – это больше, чем проблема загрузки, удовлетворяемой мощными вычислительными сервера-

ми. Гораздо большей проблемой было организовать и сделать доступным огромное количество информации в готовой к использованию форме. Это потребовало информационных систем третьего поколения. То, что в основном было двусторонним взаимодействием между хостом и клиентом, стало трехсторонним – с системой ИП в середине.

Клиенты обслуживаются хостами, связь облегчена с помощью систем ИП. Однако современные системы ИП являются большим, чем просто организатором сетевых ссылок. Они моделируют выбор и предпочтения пользователей для лучшего их обслуживания. Мы утверждаем, что триада – клиент, система ИП, хост – привлекательнее, чем сумма отдельных частей. Связь между этими тремя частями игнорируется действующей архитектурой сетевого обслуживания. Здесь мы делаем предложение, которое будет эксплуатировать эту связь для лучшего освещения некоторых аспектов использования сетевого обслуживания.

Разработчики сети ищут контент на страницах на основе информации, предоставленной владельцем сайта. Контент в сетевом сайте изначально организован на

\* Перевод Chowdary C. R., Singh A. K., Nelakanti A. Responding to retrieval: A proposal to use retrieval information for better presentation of website content.— <http://arxiv.org/pdf/1501.04509v1.pdf>

\*\* <http://www.internetlivestats.com/total-number-of-websites/>  
27/10/2014

основе категоризации информации и соответственно располагается разработчиком. В этом полном процессе современной парадигмы разработки *запрос* не играет роли во время разработки или фаз презентации сетевого сайта. Но, когда поисковый запрос задан системе ИП, то она ищет ссылки готовых сетевых страниц без учета запроса *со стороны хоста*. Так происходит потому, что поиск и управление контентом считаются взаимоключающими, т.е. система управления контентом не знает о поисковой системе, а поисковая система не знает, как контент-провайдер может ответить на запрос. Из-за этого недостатка и контент-провайдер, и система ИП находятся в процессе разработки (реализации). В этой статье мы пытаемся изучить этот вопрос, предлагая архитектуру, дающую возможность серверу множества сетевых сайтов представлять контент, основанный на запросе, изложенном пользователем.

### Предлагаемая архитектура

Схема предлагаемой нам архитектуры представлена на рисунке. Этот сценарий состоит в том, что пользователь обращается к поисковой системе и делает запрос. Поисковая система представляет поисковые результаты пользователю. Из них пользователь выбирает один и направляется к целевому сетевому сайту. Когда пользователь обращается к этому сетевому сайту, поисковая система также обменивается некоторой информацией о пользователе и запросе (о требованиях конфиденциальности см. в разделе Неприкосновенность и персонализация) с сервером, обслуживающим сетевой сайт. Хост сервер использует эту информацию для представления контента таким образом, чтобы пользователь имел лучший поисковый опыт (см. раздел Представление контента на основе запроса). Это представление, например, может облегчить пользователю нахождение определенных вещей. Хост сервер в таком случае будет обеспечивать обратную связь с поисковой системой (снова возникает вопрос о требованиях конфиденциальности) на основе продолжительности пребывания пользователя на данном сетевом сайте и его активности в это время. Так как информация, обмениваемая с хост сервером, анонимна, то имеет место обратная связь с учетом поисковой системы. Поисковая система в таком случае будет использовать эту обратную связь для выдачи лучших результатов в будущем (см. раздел Поиск на основе обратной связи). Конечный результат будет служить лучшей синхронизацией между поисковой системой и хост сервером в целях представления более совершенных результатов пользователю. Анонимность, опция отказа от услуги и персонализация будут центральными требованиями, усиленными протоколом (см. раздел Протокол поискового ответа), чтобы предотвратить любое неправильное обращение, которое может быть вызвано обменом информацией.

### Представление контента на основе запроса

Современные сетевые серверы не принимают во внимание запрос при представлении контента пользователю. Сообщалось о множестве работ по улучшению архитектуры сетевых серверов для различных применений [1,6,4]. Доступны многие модели для сравнения архитектур серверов [3]. Работа [7] обсуждает улучшения эффективности сетевых сайтов путем использования конечных серверов в архитектуре Fog Computing Architecture. По нашим сведениям не было попыток исполь-

зовать запрос к хост серверу для представления контента пользователю.

Если хост сервер может представить контент пользователю, основываясь на запросе, тогда он будет полезен и пользователю, и головной организации. Предположим, что пользователь *A* задает запрос «популярные фильмы в жанре экшен + старые», а пользователь *B* – «популярные фильмы в жанре экшен + самые новые». Допустим, что оба пользователя получили *Ссылку 1* в качестве своей первой ссылки. Мы предполагаем, что хост сервер *Ссылки 1* должен представлять различные контенты для каждого из пользователей на основе их запроса. В таком случае хост сервер может представлять список старых фильмов жанра экшен (возможно с других страниц хост сервера) для пользователя *A* и список новых фильмов в стиле экшен для *B*, в обоих случаях помимо контента на *Ссылку 1*.

Для получения максимальной выгоды от такого рода архитектуры современные системы управления контентом (CMS), подобные Drupal, Joomla, Django и т.д., должны быть переконструированы, чтобы учитывать запросы пользователей в представлении конечной сетевой страницы, предъявляемой пользователю. Это позволит хост серверу (и CMS) играть активную роль в процессе поиска контента. Так как контент-провайдер знает о контенте гораздо больше, чем поисковая система, все это знание может быть использовано для представления пользователю динамичного контента на основе запроса.

### Поиск на основе обратной связи

Классическая или упрощенная поисковая система [2] только учитывает запрос для поиска. Некоторые современные поисковые системы идут дальше и используют доступную информацию о пользователе для персонализации результатов. Однако они не принимают во внимание активность пользователя, когда он однажды сделал выбор и зашел на одну из сетевых страниц. В предлагаемой архитектуре анонимная информация об активности пользователя будет доступной для поисковой системы. Таким образом станет возможным разработать алгоритмы, учитывающие эту активность. Некоторая работа в этом направлении была предложена в [5].

Детали этой активности могут включать информацию, такую как другие ссылки на сетевой сайт, к которому обратился (кликнул) пользователь, и общее время, проведенное пользователем на этом сетевом сайте и на различных страницах. Поисковая система, осознающая эту обратную связь от хост сервера, должна интуитивно функционировать лучше.

Некоторые современные поисковые системы также предоставляют дополнительные ссылки как часть резюме «сниппета» при представлении результатов поиска пользователю. Такие сниппеты могут быть лучше подготовлены с предполагаемой обратной связью от хост сервера.

Кроме того, и что важно, хост сервер может предоставлять дополнительную информацию о своем контенте как часть обратной связи. Такая дополнительная информация будет основана на запросе и знании контента, доступного для хост сервера. Это позволит контент-провайдеру проявить себя в представлении сниппета для определенного сетевого сайта. Поисковая система может использовать или нет эту информацию в зависимости от алгоритма подготовки поиска и сниппета.



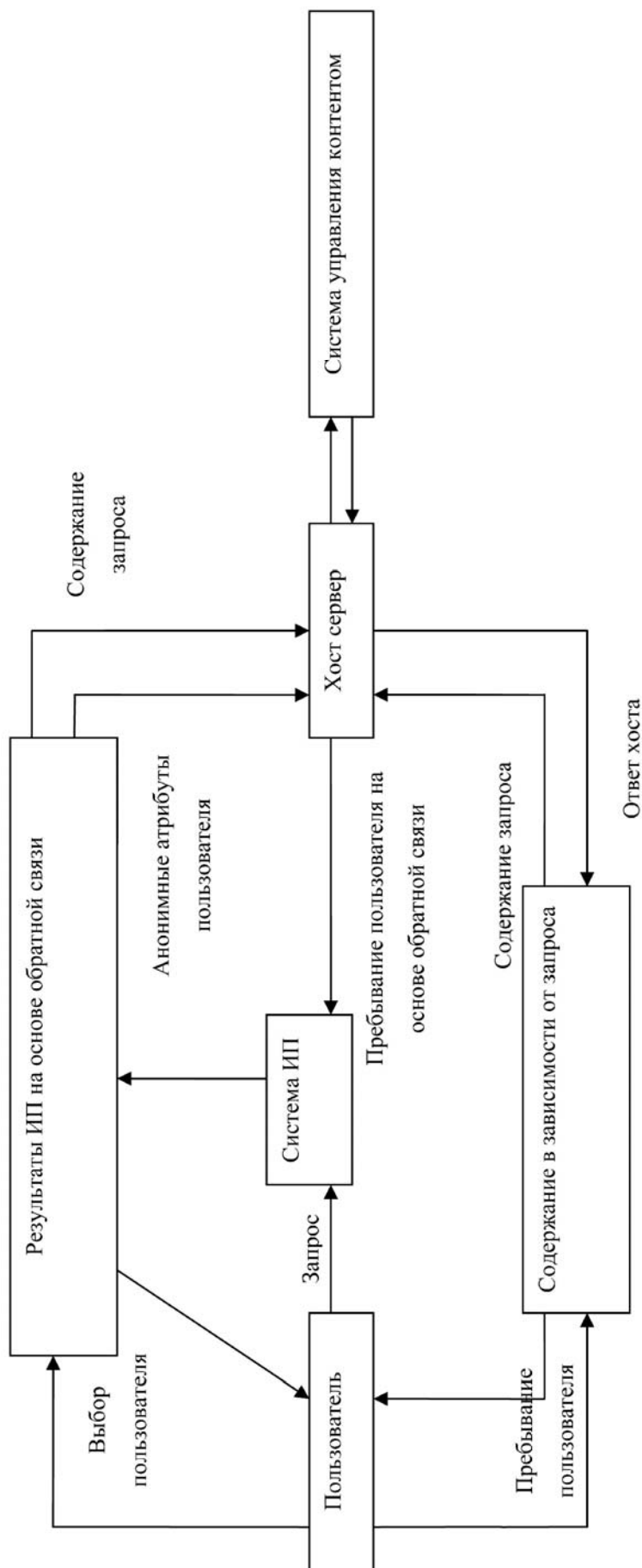


Рис. Предлагаемая архитектура для более глубокого ИП и систем управления контентом (CMS)

## Конфиденциальность и персонализация

Наше предложение требует, чтобы поисковая система обменивалась некоторой информацией о пользователе и запросе с хост сервером. Оно также требует, чтобы хост сервер обеспечивал обратную связь с поисковой системой на основе пребывания пользователя на сетевом сайте после его выбора пользователем из поисковых результатов. Этот дополнительный обмен информацией непосредственно ставит вопросы конфиденциальности. Если наше предложение применяется, его подробная версия должна включать строгие требования для изучения всех возможных проблем конфиденциальности. Приведем ниже некоторые из этих требований:

- Первым таким требованием является то, что личность пользователя, даже если она известна поисковой системе, не будет открыта хост серверу. Какая бы то ни было информация, которой обмениваются с хост сервером, должна быть строго анонимной, чтобы не раскрывать личность пользователя.

- Вторым требованием является то, что только релевантная информация будет предметом обмена. Если мы посмотрим на эту информацию как на список пар значимых атрибутов, то тогда только это подмножество пар значимых атрибутов будет представлять предмет обмена с хост сервером, что необходимо знать хост серверу с целью лучшего представления его контента.

- Третьим требованием служит то, что опция отказа от услуги будет доступна как для пользователя, так и для хост сервера. Пользователь будет знать об обмене информацией и решать, допустить или нет такой обмен. Информация будет предметом обмена, только в случае явного согласия пользователя. При отсутствии такого обмена не будет. Подобным образом хост сервер будет решать, обеспечивать ли обратную связь с поисковой системой или нет, и в случае отказа произойдет невыполнение обязательств.

- Четвертое требование заключается в том, что и пользователь, и хост сервер должны иметь возможность переделывать (модифицировать) обмен информацией. Если они решают обмениваться информацией, то в дальнейшем будут наделены опцией отбора определенных атрибутов, которыми они хотят обмениваться. Например, если поисковая система знает о местонахождении пользователя, его возрасте, гендерной принадлежности и языке, тогда пользователь может принять решение об обмене сведениями только о местонахождении и языке.

- Пользователь должен будет быть проинформирован, что деятельность на посещаемом сетевом сайте может быть использована для предоставления обратной связи поисковой системе. И тогда пользователь будет решать, какая часть активности на сетевом сайте может быть использована для предоставления обратной связи поисковой системе.

Если наше предложение будет детально проработано в будущем, то можно определить больше таких требований и изучить их.

Даже после изучения этих вопросов одна проблема все еще остается, это касается предложенной архитектуры. Даже если подлежащая обмену информация анонимна, и хост сервер не знает личность пользователя, поисковая система может ее знать и иметь возможность связать активность пользователя на посещаемом сетевом сайте с его личностью. Возникает

вопрос, должна ли поисковая система получать больше знания о пользователе, чем гарантировано. Это проблематичный этический вопрос, требующий дальнейшего исследования.

## Протокол поискового ответа

Существует много различных видов поисковых систем. Так же имеется много различных типов хост серверов и систем управления контентом. Если между ними должен быть поток информации, как говорится в предыдущих разделах, то он должен четко регулироваться, чтобы можно было применять системы без какого-либо конфликта. Это требует точно определенного и хорошо разработанного протокола. Назовем его Протоколом поискового ответа (Retrieval Response Protocol – RRP).

Протокол поискового ответа будет регулировать поток информации между поисковой системой и хост сервером. Данный протокол будет использоваться для инициации, поддержки и закрытия поисковой сессии. Как только пользователь выберет один результат из результатов, представленных поисковой системой в ответ на его запрос, поисковая сессия будет открыта. Завершением сессии, вероятно, будет перерыв в работе, основанный на том, что нет иного способа узнать, когда пользователь покинул сетевой сайт.

Во время действия сессии поисковая система будет первой обмениваться информацией о пользователе и запросе с хост сервером. После этого, основываясь на активности пользователя, хост сервер обеспечит обратную связь с поисковой системой. Вся деятельность во время данной сессии будет предметом требований относительно конфиденциальности и персонализации, и проект протокола должен это учитывать.

Протокол должен быть разработан в целях регулирования этой поисковой сессией. Мы оставляем разработку этого протокола на будущее.

## Заключение

В современной парадигме информационного поиска хост сервер не использует информацию запроса для представления контента. Поисковая система не знает, что происходит после того, как пользователь выбрал результат поиска. А хост также не имеет доступа к информации, доступной поисковой системе. Мы представили схему архитектуры, которая изучает эти вопросы. Цель – обеспечить лучший поисковый опыт пользователю через лучшее представление контента на основе запроса и лучшие поисковые результаты на основе обратной связи поисковой системой с хост сервером. Поисковая система будет обмениваться некоторой информацией с хост сервером, а хост сервер, в свою очередь, обеспечит релевантную обратную связь с поисковой системой на основе пребывания пользователя на сетевом сайте. Хост использует всю связанную с запросом информацию для динамичного представления контента. Эта пересмотренная парадигма для информационного поиска также затрагивает вопросы конфиденциальности, которые должны быть тщательно изучены. Она также требует нового протокола для контента поискового ответа, который мы кратко описали. Этот протокол будет регулировать поток информации между поисковой системой и хост сервером, подчиняясь требованиям конфиденциальности и персонализации.

## Литература

1. *Begen A., Akgul T., Baugher M.* Watching video over the web: Part 1: Streaming protocols// IEEE Internet Computing.— 2011.— Vol. 15, No.2 (March).— P. 54–63.
2. *Brin S., Page L.* The anatomy of a large-scale hypertextual web search engine// Proceedings of the Seventh International Conference on World Wide Web 7, WWW7, pages 107–117. Amsterdam, The Netherlands, The Netherlands, 1998.— Elsevier Science Publishers B. V., 1998.
3. *Harji A.S., Bubr P.A., Brecht T.* Comparing high-performance multi-core web-server architectures// Proceedings of the 5th Annual International Systems and Storage Conference, SYSTOR '12, pages 1:1–1:12, New York, NY, USA, 2012. — ACM.
4. *Hashemian R., Krishnamurthy D., Arlitt M., Carlson N.* Improving the scalability of a multi-core web server// Proceedings of the 4th ACM/SPEC International Conference on Performance Engineering, ICPE '13, pages 161–172, New York, NY, USA, 2013. — ACM.
5. *Shen X., Tan B., Zhai C.X.* Context-sensitive information retrieval using implicit feedback// Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '05, pages 43–50, New York, NY, USA, 2005. —ACM.
6. *Veal B., Foong A.* Performance scalability of a multi-core web server// Proceedings of the 3<sup>rd</sup> ACM/IEEE Symposium on Architecture for Networking and Communications Systems, ANCS '07, pages 57–66, New York, NY, USA, 2007. — ACM.
7. *Zhu J., Chan D.S., Prabhu M.S., Natarajan P., Hu H., Bonomi F.* Improving web sites performance using edge servers in fog computing architecture// IEEE 7th International Symposium on Service Oriented System Engineering (SOSE), pages 320–323, March 2013.

# Отчет о Международной конференции «Направления в издательстве и контексты 2014 г. — цифровые авторы и электронные книги», 8-9 декабря 2014 г., г. Пула, Хорватия\*

Елена МАЦЕВИЧЮТЕ  
(Elena MACEVIČIŪTĒ)

Шведская школа библиотековедения и информатики, Буросский университет, г. Бурос, Швеция

## Введение

Вторая международная конференция по современности была организована Отделением информационных наук в университетах городов Задар и Осиек во время фестиваля книги и авторов в г. Пула. Соучредителями и спонсорами конференции были ассоциация Sa(n)jam Knjige Iстрия и министерства культуры и образования (науки) и спорта Республики Хорватия. Они планировали, что конференция даст лучшее понимание процессов, которые в настоящее время меняют книжную торговлю и чтение в Европе.

За два дня конференции 26 представителей из 9 стран обсудили вопросы книжного рынка, читательского спроса, новых моделей издательства, состояние исследований в издательстве и обучении, вызовы, брошенные электронными книгами авторам и другим участникам современной системы коммуникации. Участники представили 22 доклада по этим темам, но наиболее успешной частью конференции была активная дискуссия, вызванная этими докладами и поднятыми в них вопросами. Обсуждение проходило во время презентаций докладов в конце каждого дня или в течение перерывов. Студенты хорватских университетов слушали обсуждения и принимали в них участие. Атмосфера конференции была дружелюбной, и все участники не только были заинтересованы обсуждаемыми темами, но и проявляли широкую осведомленность. Интеллектуальный уровень был высоким, а интересы — разносторонними.

\* Перевод Macevičūtē E. Conference report: International conference “Publishing trends and contexts 2014, focus :Digital authors and electronic books”, 8-9 December, 2014, Pula, Croatia. — <http://informationr.net/ir/20-1/paper653.html>

## Поиск концептуальных основ

Некоторые докладчики искали базовые **концептуальные основы**, которые могли бы помочь объяснить им сущность современного издательства. Презентация Майкла Бхаскара (издательство Profile Books, Великобритания) основывалась на его книге «The content machine». Он подчеркивал, что теория контента может помочь в понимании изменений в издательстве. Создание контента управляло процессом распространения, а также отображало, как мы читаем тексты и насколько, по опыту, контент субъективно воспринимаем. Издание всегда некоторым образом расширяет контент с помощью определенной структуры. Изучение таких структур — от объективной стороны распространения до субъективной стороны опыта — будет усиливать наше знание относительно издательства и чтения.

Адриан ван дер Вил (Лейденский университет, Нидерланды) проложил путь к презентации Бхаскара в самом начале конференции, так как изучал культурно значимые различия между печатными и цифровыми книгами и в результате снизил важность книжной культуры, но поднял значимость культуры чтения. Книга с ее собственными наследуемыми технологическими свойствами может служить в качестве среды свободного отвращения для обдумывания, а также как память и социальная технология. Изменяющиеся технологические свойства не ликвидируют старые средства, а меняют способ, с помощью которого люди используют и создают тексты, на способы, связанные с их культурно значимыми системами.

Потребность в концептуализации изменения, происходящего в области постоянно совершенствующейся практики издательства, стимулирует появление исследований, связанных с издательством. Одно из заседаний было посвящено изучению корней исследований издательства, некоторые другие доклады также изучали этот

вопрос. Софи Ноэль (Сорбонна, Франция) исследовала происхождение этого трудного научного объекта. Она предположила, что сегодня пока еще не существует единой области исследования издательства. Она исследуется историками, экономистами сферы культуры, литературоведами, социологами и учеными в области информации и коммуникации. Симптоматическое появление термина *исследование издательства* означает потребность в профессиональной практике, а также в суммировании и обобщении растущего объема междисциплинарного исследования. Многим прикладным дисциплинам трудно сопротивляться экономическим давлению, чтобы обслуживать потребности практики, они остаются узкими и с научной точки зрения, поэтому важно критически подойти и оценить их легитимность и основы, а также найти концептуальное согласие.

Бертран Лежандр (Сорбонна, Франция) изучил классические работы, лежащие в основе исследований издательства, особый акцент сделан на французских классиках, таких как Люсьен Фебвр, Анри-Жан Мартин, Роже Шартье и других книгах историков. Он также отметил, что издательское дело является частью культурной индустрии и, таким образом, изучается в различных дисциплинах, например, экономистом сферы культуры Франсуазой Бенаму, социологами проблем чтения Кристофом Эвансом, Оливье Донном и другими. Многие проекты, нацеленные на стимулирование чтения и изучение ситуации в издательском деле, поддерживаются государством. Они (проекты) поддерживают критический подход ко всем частям издательского цикла.

По мнению Кристофа Блэзи (Университет Йоганна Гутенберга, г. Майнц, Германия), подобные вопросы и проблемы легитимности, междисциплинарности, положения в университетском и научном сообществах изучаются по опыту современными исследованиями относительно книг и издательства в Германии. Он также провел опрос ряда ученых в британских университетах. Они подтвердили проблему определения исследования этой области, отсутствие оригинальных теоретических и методологических подходов. Область тесно связана с издательским делом и более применима, чем теория, но также изучает вопросы, которые не имеют отношения к этой индустрии, такие как грамотность и чтение, культурное преобладание англо-американского производства и т.д. Можно сделать вывод, что область исследования издательского дела слишком мала, чтобы называться дисциплиной. Она оригинальна и отличается от исследований медиа, так как не фокусируется на вопросах производительности. Она развивает некоторые черты молодой дисциплины, но пока позволяет много свободы тем, кто в ней работает.

Исследование Франьо Пегара (Задарский университет, Хорватия) и Зорана Велагича (Осиекский университет, Хорватия) стоит несколько в стороне от остальной части докладов на эту тему. Они предложили библиометрическое исследование изучений издательского дела, используя определенную методологию восьми *ядерных* монографий, размещенных в библиографической базе данных. Выбор монографий и сам метод критически обсуждались аудиторией, но все согласились, что результаты этого исследования, представляющего различные части изучения издательского дела, отношения между ними и связи ссылок между работающими в этой области учеными были очень интересными и заслуживающими того, чтобы более широко применить библиометрию в этой области. Дисциплина показывает поведенческую характеристику области гуманитарных

наук с большими ссылками на книги, чем на статьи (758 и 334 соответственно).

Вообще эта концептуальная часть напомнила мне обсуждения, которые мы наблюдали в библиотекведении и информационной науке, в области ухода за больными и информатике и многих других прикладных дисциплинах, появляющихся в результате потребностей обслуживания профессиональных видов деятельности. Проблемы, недостатки и преимущества кажутся одинаковыми по всему спектру этих дисциплин. Единственное, что меня больше всего волновало, так это проблема научных возможностей, поскольку она относится к экономическим ресурсам, необходимым для проведения исследования. Так как профессиональные практики больше полагаются на консультирование, чем на исследование, то маловероятно, что индустрия обеспечит эти ресурсы для области научного исследования. Соперничество с авторитетными и опытными научными игроками за скудные научные ресурсы, доступные через государственное финансирование и государственные фонды, требует научных возможностей и признания со стороны других научных участников. Любая стратегия достижения этих ресурсов (финансирования) должна приветствоваться.

### Программы обучения

Одна стратегия, а именно, учреждение научных образовательных программ в университетах изучалась в рамках специальной сессии. Интересно отметить, что **образование в области издательского дела** было представлено тремя сравнительно новыми для этой области лицами, хотя все они могут похвастаться прочной традицией изучения книг. Все трое относятся к центральной части Восточной Европы: Литва, Польша и Хорватия. Аушра Навискене (Вильнюсский университет, Литва) рассказала о развитии исследований в области издательского дела в аспирантурах, представляющих основы образования третьего уровня в соответствующих областях в Вильнюсском университете как единственном научном учреждении в Литве, в котором можно получить степень в области информации и коммуникации. Это общий контекст, в котором ученые, имеющие степень в области издательского дела, развиваются в этой небольшой стране с богатой традицией исследования книг и образования. Вроцлавский университет (Польша) имеет одинаково богатые традиции в обеих областях. Ева Яблоньска-Стефанович познакомила с подъемом образования в области издательского дела во Вроцлавском университете, осветив основные проблемы создателей этой сферы, характерные черты образования и его статус и другие родственные университетские программы. Энтузиазм преподавателей и удовлетворение студентов были осязаемой особенностью ее презентации. Йосипа Стетхофер (Осиекский университет) подняла вопрос подхода к установлению образовательной программы в издательском деле. Она сделала обзор существующих образовательных программ и курсов в США и Великобритании и затронула вопрос, могут ли они служить моделью для подобной программы в Хорватии. Большинство преподавателей, осуществляющих эти программы, приходит из издательской индустрии, могут ли они обеспечить научную подготовку, требуемую в современной меняющейся среде? Это было своего рода антитезой и комплиментом для предыдущего доклада, утверждающего существование основных навыков, которые должны обеспечиваться этими программами независимо от времени.

## Эмпирическое исследование

Участники конференции рассмотрели не только научные и образовательные вопросы исследования издательского дела, но и представили результаты недавнего **эмпирического исследования**, касающегося каждого аспекта книжного цикла. Том Уилсон (Буросский университет, Швеция) говорил о дилемме *издателей*, очевидной из недавних исследований мнений издателей относительно электронных книг в Швеции, Литве и Хорватии. Эти три страны были выбраны как представители рынков «малого языка», работающих иначе, чем международные рынки английского или испанского языков. Докладчик сделал акцент на амбивалентности и неуверенности, ощущаемых издателями в отношении электронных книг и изменений, которые они вносят. С одной стороны, они ожидают, что изменится немного (например, их позиция на рынке, роли книгохранилищ или библиотек), а с другой – издатели планируют и претворяют изменения в свою собственную деятельность, такую как развитие своих самоиздательских платформ или продажа книг через собственные сетевые сайты. Несмотря на общее ожидание небольшого изменения в росте рынка электронной книги в этих трех странах, большая часть издателей планирует производство электронных книг в следующие пять лет.

Ангус Филипс (Оксфордский международный центр по исследованиям издательского дела, Великобритания) показал, что амбивалентность представлена не только на рынках *малого языка*, но также и по всему миру. Он предположил, что издатели нуждаются в строительстве того, что мы называем «цифровым капиталом», т.е. в новых способах оперирования, предоставляющих услуги авторам, непосредственно доходя до конечного читателя, выработка стратегий новых брендов, соединения широкие сообщества книгоиздателей, библиотекарей и читателей для совместного создания ценности через книги и чтение. Эти виды деятельности должны объединить все виды капитала: экономический, финансовый, общественный и т.д.

Ивана Хебранг Гргич (Загребский университет, Хорватия) подошла к меняющейся среде издательства с точки зрения научной коммуникации и журнала. Она представила обзор хорватских научных журналов, надеющихся на инновационные особенности, создаваемые цифровым издательством. Она обнаружила, что эти журналы статичны и, как правило, имитируют свои печатные версии и не используют доступные цифровые возможности; согласно Ангусу Филипсу, цифровой капитал пока не является характерной особенностью издательства научных журналов в Хорватии.

Ивана Деспот, Ивана Львак Лебеда и Нивес Томашевич (Задарский университет, Хорватия) изучили, как доступ к свободному контенту влияет на развитие новых бизнес-моделей *freemium*. Эти новые модели сочетают свободный доступ с привилегиями класса премиум, доступными за клиентскую плату. Ряд существующих моделей и примеров, таких как свободный просмотр текста за ограниченный период времени, различные особенности контента за плату и бесплатно, публикация сериями и другие, были представлены авторами. Эти случаи отражают новые модели маркетинга и совместных усилий создателей контента для достижения более широкой читательской аудитории.

Клер Скрисес была первой, кто ввел тему маркетинга, поскольку принимала во внимание способы пересмотра и усовершенствования своей ранее изданной книги «Маркетинговая литература: создание современного

произведения в Великобритании». Так как ее монография была основана на изучении случаев маркетинга книг-бестселлеров, она проверила свои идеи на новых случаях, которые могут представлять новую ситуацию в издательстве, включая случаи цифрового и самостоятельного издательства авторов бестселлеров. Один такой случай исследования литовского автора был представлен Арунасом Гудинавичюсом (Вильнюсский университет, Литва). Он внимательно изучил маркетинговые усилия Андриуса Тапинаса, который написал литовский бестселлер «Час волка», перевел его на английский язык, самостоятельно издал и выставил этот перевод на продажу через Интернет. Хотя его книга стала успешной на литовском рынке, попытка самостоятельной продажи, что требуется отметить, за пределами Литвы была весьма высокой, но не дала ожидаемых результатов.

Другой аспект самостоятельного издательства был изучен Астой Урбанавичюте (Вильнюсский университет, Литва). Она сравнила самоиздающиеся журналы по культуре советской эпохи (самиздат) и самоиздающиеся электронные тексты по культуре, распространяющиеся через социальные сети и блоги. Хотя производители современного самоиздаваемого текста не подвергают сомнению опыт ранних феноменов или самиздата, существуют некоторые общие черты между самостоятельным издательством в обоих периодах: сопротивление (политическому или коммерческому правилу), свобода от всевозможных ограничений, отсутствие внимания к качеству или литературным достоинствам текстов.

## Другие действующие лица

Помимо издателей участники конференции обсуждали вопросы, связанные с другими представителями издательского дела. Томислав Якопеч (Осиекский университет, Хорватия) представил результаты своего исследования собирателей электронных книг, включая также платформы для издания электронного контента и продавцов электронных книг. Основной целью его исследования было определить собирателя электронных книг путем изучения его действий, услуг, манер производства и других вопросов. Собиратели электронных книг представляют собой переходную структуру между авторами электронных книг и хранителями или покупателями книг, обеспечивая важные стратегии продвижения в качестве основных услуг. Елена Мацевичюте (Буросский университет) представила результаты исследования мнений продавцов книг, проведенного в Швеции. Продавцы книг еще «не вскочили в вагон с электронной книгой». Только некоторые энтузиасты продают их в реальных книжных магазинах, если у них есть собственная технология и доступ к поставщикам электронных книг. Основным препятствием для продажи электронных книг является отсутствие спроса со стороны их покупателей, которые очевидно ищут электронные книги в сети. Тем не менее, шведские продавцы книг надеются выжить в будущем и не считают электронные книги угрозой их позиции на рынке. Напротив, многие из них рассчитывают начать продавать электронные книги в будущем.

Нивес Томашевич (Задарский университет, Хорватия) рассказала о книжных ярмарках и их роли в цифровом веке. Она сравнила содержимое нескольких основных книжных ярмарок, модели посещаемости и сопоставила их с виртуальными встречами в цифровых окружающих средах. Не представляется, что реальные книжные ярмарки находятся под угрозой из-за цифрового издательства. Наоборот, они находятся среди пло-

щадок для обсуждения и проверяют возможности цифрового книжного производства, распространения и чтения. Майя Крталич (Осиекский университет, Хорватия) затронула проблему, которая очень редко обсуждается на конференциях, посвященных издательскому делу, хотя она довольно часто встречается на дискуссионных форумах библиотечарей. Она говорила о юридических, технических, общественных и организационных вопросах сохранения электронной книги и ее долговременной доступности. Также она сравнила ответственность и возможности издателей и библиотечарей в этом отношении и подняла один из важных вопросов: кто должен отвечать за долговременное хранение электронных книг и как это следует организовать? Кажется, что некоторые участники конференции никогда не задумывались над этим вопросом и впервые слышали его обсуждение.

### Исследование чтения

Последнее, но не менее важное, я хотела бы сообщить о презентациях относительно вопросов чтения. Было только две из них на этой конференции, поднятые докладчиками вопросы заслуживают более пристального внимания и дальнейшего изучения. Бенуа Берту (Сорбонна, Франция) представил исследование цифрового чтения комиксов во Франции и сравнил его в некоторой степени со статистикой чтения, представленной «Евробарометром» («Евробарометр» - серия регулярных сравнительных исследований общественного мнения в странах, входящих в Европейский союз. Проводится Европейской комиссией два раза в год (весной и осенью), начиная с 1973 г. Тематика опросов периодически меняется. — *Прим. ред.*). Одно из различий состоит в том, что до сих пор мужчины образуют большинство читателей комиксов, в то время как в других видах чтения доминируют женщины. Но число женщин, читающих комиксы, растет. Читатели ищут совета не у профессионалов, а преимущественно у своих коллег. Большинство читателей находят комиксы через библиотеки, и только половина тех, кто имеет доступ к ним через Интернет, платит за материал. Я хочу подвергнуть сомнению утверждение автора, что все, кто не платит, используют пиратские материалы; число бесплатных, но полностью законных сайтов с комиксами

вполне значительно, даже если не считать те, что доступны через бесплатные газеты. Одним из интересных аспектов этой презентации, заслуживающим дальнейшего изучения, является доступ к комиксам и отношение к использованию других визуальных материалов, таких как фильмы, видеоролики и картинки. Миха Ковач (Люблянский университет, Словения) представил самое интересное исследование по распространению книг и чтения в Словении. Он показал, что под прикрытием стабильности статистики покупки книг почти за 40 лет (с 1973 г.) происходит распад словенского книжного рынка. Этот сдвиг усиливается предпочтением публики читать более дешевый или свободно доступный через библиотеки материал, загружать его из Интернета или покупать в легко доступных местах, таких как супермаркеты. Это приводит к исчезновению домашних библиотек. Число читателей, читающих по-словенски и, по крайней мере, на одном другом иностранном языке, растет, и появляется группа читателей (4%), читающих только английскую литературу, использующих английский контент и общающихся в режиме онлайн на английском языке. Молодое поколение, покидающее словенский культурный контекст, демонстрирует не только падение традиционного культурного сообщества, но и разрушение воображаемых национальных сообществ.

### Заключение

Я бы не хотела завершать обзор на этой пессимистичной ноте. Каждое действие вызывает некоторую реакцию. Когда мы видим результаты реакции, то стоит узнать, что ее вызвало. Изменения в целых моделях чтения не происходят сами по себе, они являются звеном в цепи изменений. Ранняя диагностика (а мы находимся на ранних стадиях изменений не только в чтении, но и в создании текстов) может помочь в нахождении причин и устранении недугов. Мы можем также считать, что эти изменения, вызванные другими более позитивными процессами в наших языках и культурах, хотя и незначительные, но настолько же важны, как и международные разработки.

Некоторые презентации, о которых здесь говорилось, будут напечатаны в электронном журнале *Libellarium*.

# Библиометрически улучшенный информационный поиск: 2-й международный семинар BIR\*

**Филипп МАИР**  
(Philipp MAYR),

**Андреа ШАРНХОРСТ**  
(Andrea SCHARNHORST),

**Петер МУТШКЕ**  
(Peter MUTSCHKE)

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

**Инго ФРОМХОЛЬЦ**  
(Ingo FROMMHOLZ)

Отделение вычислительной науки  
и техники, Бедфордширский университет,  
г. Лутон, Великобритания

*Семинар BIR (Bibliometric-enhanced Information Retrieval) объединяет вместе экспертов сообществ, которые часто воспринимаются как отдельные: библиометрия/ наукометрия/ информетрия с одной стороны и информационный поиск – с другой. Наша мотивация как организаторов данного семинара началась с наблюдения того, что основные дискурсы в обеих областях различны, что сообщества только частично перекрывают друг друга и с веры в то, что передача знаний будет полезна обеим сторонам. Библиометрические методы еще не применяются широко для улучшения процессов поиска в цифровых библиотеках, хотя они предоставляют пользователям результаты с дополнительными функциональными возможностями. С другой стороны, все больше информационных специалистов, работающих в библиотеках и архивах, сопротивляются применению библиометрических методов в своих службах. В таком случае обмен знаниями становится более важным. Первый семинар определил научную повестку дня – знакомство с другими методами, сообщения о текущих научных проблемах и широкое обсуждение общих интересов. Последующий семинар продолжает всестороннее общение, но также выделяет одну центральную проблему. В частности, мы будем изучать, как статистическое моделирование грамотности может улучшить поисковые службы для определенных сообществ, а также для крупных, междисциплинарных массивов, таких как Mendeley или ResearchGate. Второй семинар BIR продолжает повышать осведомленность относительно отсутствия связи между информационным поиском и библиометрией и вносит вклад в создание общей основы для включения библиометрически улучшенных служб в поиск в интерфейсе научно-поисковой системы.*

## Введение

Информационный поиск и библиометрия уходят своими корнями далеко назад. Многие основоположники библиометрии действительно пришли в нее из области информационного поиска, которая является одной из традиционных ветвей информатики (см. например, [1]). Информационный поиск как технология служит истоком любого наукометрического\*\* изучения, и

поэтому информационный поиск относится к сфере навыков для любого представителя библиометрии/наукометрии. При использовании в оценках библиометрические методы сохраняют и снижают свою позицию в зависимости от надежности определяемых направлений работы в области или организации. При использовании в поиске информации в больших массивах эти библиометрические методы могут помочь направить внимание пользователя на возможное ядро информации в более широком поисковом корпусе знания.

Однако информационный поиск и библиометрия как особые научные области также отдалились друг от друга за последние десятилетия; при сегодняшних документальных массивах «больших данных», объединяющих аспекты краудсорсинга, рекомендации, интерактивного поиска и социальных сетей, наблюдается повышенный интерес к пересмотру информационного поиска и библиометрии в целях обеспечения передовых решений, которые помогают удовлетворять сложные, раз-

\* Перевод Mayr P., Frommholz I., Scharnhorst A., Mutschke P. Bibliometric-enhanced information retrieval: 2<sup>nd</sup> International BIR workshop.— 2015. — <http://arxiv.org/ftp/arxiv/papers/1501/1501.02646.pdf>

\*\* Слова библиометрия и наукометрия, иногда даже информетрия используются альтернативно и являются взаимозаменяемыми, наукометрия, как правило, шире и включает также исследования затрат, образования и учреждений, короче говоря, всех метрик и показателей, встречающихся в количественных исследованиях научной системы.



нообразные и долговременные информационные потребности ищущих информацию (пользователей). Это проявляется в хорошо посещаемых комбинированных текущих семинарах, таких как «Вычислительная наукометрия» (в рамках iConference 2013 и СИКМ 2013), «Сочетание библиометрии и информационного поиска» (ISSI conference 2013, <http://www.gesis.org/en/events/events-archive/conferences/issiwkshop2013/>) и семинар прошлого года ECIR BIR. Очевиден рост осознанности того, что изучение связей между библиометрическими методами и информационным поиском полезно для обоих сообществ [2, 3]. Эти семинары выявили, что дополнительная предстоящая работа в этом направлении зависит от продолжающегося роста осознанности в обоих сообществах, проявляющейся в конкретных экспериментах/изучении в существующих поисковых системах.

Также растет важность сочетания библиометрии и информационного поиска в реальных применениях [4], например относительно современного отслеживания разработок в области. Другим примером является предоставление услуг, поддерживающих ученых на уровне новейших требований их области посредством рекомендации и интерактивного поиска, например, в «инструментальных средствах для ученых», таких как Mendeley/ResearchGate или поисковых системах как Google Scholar, использующих крупные библиометрические массивы. Мы надеемся, что данный семинар внесет свой вклад в определение и дальнейшее изучение применений и решений, объединяющих оба сообщества. Первый семинар библиометрически улучшенного информационного поиска (BIR) (<http://www.gesis.org/en/events/events-archive/conferences/ecirworkshop2014/>) на конференции ECIR 2014 [5] привлек более 40 участников (главным образом из академической сферы) и вылился в три интерактивные сессии докладов [6] с живыми дискуссиями и дальнейшими действиями. Мы будем опираться на этот опыт при проведении семинара BIR 2015 (<http://www.gesis.org/en/events/events-archive/conferences/ecirworkshop2015/>). Между тем специальный выпуск журнала *Scientometrics* по «Сочетанию библиометрии и информационного поиска» под редакцией Филиппа Мапра и Андреа Шарнхорста [7] объединяет восемь статей экспертов из библиометрии/наукометрии/информетрии с одной стороны и информационного поиска – с другой.

## Цели, задачи и результаты

Предложение нашего семинара нацелено на контактирование с сообществом информационного поиска по вопросу возможных связей с библиометрией и комплексной теорией сетей, которая также изучает сети научной коммуникации. Библиометрические методы еще не широко используются в целях улучшения информационных процессов в цифровых библиотеках, но они предлагают пользователям результаты с дополнительными функциональными возможностями [8]. В качестве примера, недавние подходы показывают возможности альтернативных методов ранжирования, основанных на анализе ссылок, ведущих к улучшенному информационному поиску. Наши интересы включают поиск информации, информационный поиск, научное моделирование, сетевой анализ и цифровые библиотеки. Цель – применить способности библиометрии, наукометрии и информетрии к конкретным практическим проблемам информационного поиска и просмотра информации. Этот семинар тесно связан с семинаром BIR на конференции ECIR 2014 и стремится объединить вклады основных представителей библиометрии и ведущих специалистов информационного поиска, но выбрав тех, которые уже работают над интерфейсом между наукометрией и информационным поиском. На работе этого семинара мы больше концентрируемся на реальном экспериментировании (включая демонстрационные версии) и участии промышленности.

- Как мы можем построить системы научной информации, которые явно используют библиометрические измерения в интерфейсе пользователя?

- Как могут модели науки быть взаимосвязаны с научным, ориентированным на задачи поиском?

- Как сочетать классический информационный поиск (с акцентом на полноту и слабые связи) с более строгими библиометрическими рекомендациями?

- Как разработать схемы оценки не попав в ловушку слишком дорогостоящей структуры крупномасштабного эксперимента?

- Как сочетать разработанные в библиометрии средства, такие как CitNetExplorer или Science of Science (Sci2), с информационным поиском?

- И наоборот, может ли понимание поиска также улучшить и сами соответствующие статистические модели?

## Формат и структура семинара

Этот семинар откроется вдохновляющим вступительным докладом, чтобы дать старт размышлениям и обсуждению темы семинара. Затем последуют презентации докладов в формате, имевшем успех на EuroHCIR 2013 и 2014: каждый доклад представляется в виде 10-минутного краткого сообщения и затем обсуждается в течение 20 минут группами участников семинара, после чего следуют минутные высказывания по основным обсуждавшимся вопросам и полученным урокам. Семинар завершится круглым столом с обсуждением того, как добиться прогресса в улучшении информационного поиска с помощью библиометрических методов.

## Аудитория

Аудитории (или клиенты) информационного поиска и библиометрии различны. Традиционный информационный поиск служит отдельным информационным потребностям, и, следовательно, встроены в библиотеки, архивы и подобные собрания. Наукометрия, а с ней библиометрические методы, созрела для обслуживания научной политики. Мы предлагаем семинар на неполный рабочий день, который объединит ученых информационного поиска и цифровых библиотек по их интересу к библиометрически улучшенным подходам. Наши интересы включают поиск информации, информационный поиск, научное моделирование, сетевой анализ, и цифровые библиотеки. Цель – применить идеи библиометрии, наукометрии и информетрии к конкретным практическим проблемам информационного поиска и просмотра информации. Этот семинар тесно связан с семинаром BIR на конференции ECIR 2014 и стремится объединить вклады основных представителей библиометрии и ведущих специалистов информационного поиска, но выбрав тех, которые уже работают над интерфейсом между наукометрией и информационным поиском. На работе этого семинара мы больше концентрируемся на реальном экспериментировании (включая демонстрационные версии) и участии промышленности.

## Результаты

Статьи, представленные на семинаре BIR 2014, опубликованы в трудах в интерактивном режиме (<http://ceur-ws.org/Vol-1143/>). Еще один результат инициативы BIR оформился после проведения в рамках

конференции ISSI 2013 семинара по теме «Сочетание библиометрии и информационного поиска» в виде специального выпуска журнала *Scientometrics* [7]. С помощью предложенного семинара мы нацелены на подобную стратегию распространения, но сейчас ориентированного на ядро информационного поиска. Таким образом мы строим последовательность изучений, представлений, результатов, задокументированных в научном дискурсе, и сбора достаточного материала для строительства соответствующего моста между библиометрией и информационным поиском.

## Литература

1. *White H. D., McCain K. W.* Visualizing a discipline: An author co-citation analysis of information science, 1972–1995// *Journal of the American Society for Information Science*. — 1998.— Vol.49.— P. 327–355.
2. *Wolfram D.* The symbiotic relationship between information retrieval and informetrics// *Scientometrics*.— 2015.— doi:10.1007/s11192-014-1479-0
3. *Abbasi M. K., Frommholz I.* Cluster-based polyrepresentation as science modelling approach for information retrieval// *Scientometrics*.— 2015.— doi:10.1007/s11192-014-1478-1

4. *Jack K., López-García P., Hristakeva M., Kern R.* {{citation needed}}: Filling in Wikipedia's citation shaped holes// *Bibliometric-enhanced Information Retrieval, ECIR*. — Amsterdam, 2014—<http://ceur-ws.org/Vol-1143/paper6.pdf>

5. *Mayr P., Scharnhorst A., Larsen B., Schaer P., Mutschke P.* *Bibliometric-enhanced Information Retrieval*// M. et al. de Rijke (Ed.), *36th European Conference on IR Research, ECIR 2014, Amsterdam, The Netherlands, April 13-16, 2014. Proceedings* (pp. 798–801). — Springer International Publishing, 2014. doi:10.1007/978-3-319-06028-6\_99

6. *Mayr P., Schaer P., Scharnhorst A., Mutschke P.* Editorial for the *Bibliometric-enhanced Information Retrieval Workshop at ECIR 2014* //P. Mayr et al. (Ed.), *Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval* (pp. 1–4). — Amsterdam, NL, 2014.— <http://ceur-ws.org/Vol-1143/editorial.pdf>

7. *Mayr P., Scharnhorst A.* *Scientometrics and information retrieval - weak-links revitalized*// *Scientometrics*. — 2015.— doi:10.1007/s11192-014-1484-3

8. *Mutschke P., Mayr P., Schaer P., Sure Y.* *Science models as value-added services for scholarly information systems*// *Scientometrics*.— 2011.— Vol. 89, No.1.— P. 349–364. — doi:10.1007/s11192-011-0430-x

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](mailto:mfi@viniti.ru).

Ответственный за выпуск *Л. В. Кобзева*

Компьютерная верстка *М. А. Филимонова*

ИД № 04689 от 28.04.2001 г.

Подписано в печать 26.05.2015 г.

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

Усл. печ. л. 4,50 Уч.-изд. л. 5,24 Тираж 42 экз.

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

Тел. (499) 155-44-95

**Федеральное государственное бюджетное учреждение науки  
ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ  
ИНФОРМАЦИИ РОССИЙСКОЙ АКАДЕМИИ НАУК**

**предлагает научным работникам, аспирантам и другим специалистам в области естественных, точных и технических наук, желающим быстро и эффективно опубликовать результаты своей научной и научно-производственной деятельности, использовать способ публикации своих работ через систему депонирования.**

«Депонирование (передача на хранение) – особый метод публикации научных работ (отдельных статей, обзоров, монографий, сборников научных трудов, материалов научных конференций, симпозиумов, съездов, семинаров) узкоспециального профиля, разрешенных в установленном порядке к открытому опубликованию, широкое тиражирование которых, как правило, в силу их узкой специализации, не считается целесообразным, а также работ широкого профиля, срочная информация о которых необходима для утверждения их приоритета. Депонирование предусматривает прием, учет, регистрацию, хранение научных работ и обязательное размещение информации о них в специальных информационных изданиях».

Подготовка и передача на депонирование научных работ происходит в соответствии с «Инструкцией о порядке депонирования научных работ по естественным, техническим, социальным и гуманитарным наукам» (М., 2013).

Депонированные научные работы находятся на хранении в депозитарии ВИНТИ РАН, копии работ предоставляются заинтересованным организациям и специалистам на бумажном и электронном носителях и являются официальной публикацией.

Информация о депонированных научных работах включается в информационные издания ВИНТИ РАН, в РЖ ВИНТИ РАН и БД ВИНТИ РАН и аннотированный библиографический указатель «Депонированные научные работы».

Подать научную работу на депонирование можно, обратившись в Отдел депонирования ВИНТИ РАН по адресу:

**125190, Москва, ул. Усиевича, 20.**

**ВИНТИ РАН, Отдел депонирования научных работ.**

**Тел.: 8 (499) 155-43-28, Факс: 8 (499) 943-00-60.**

**e-mail: [dep@viniti.ru](mailto:dep@viniti.ru)**

С инструкцией о порядке депонирования можно ознакомиться на сайте ВИНТИ РАН: <http://www.viniti.ru>