

# МЕЖДУНАРОДНЫЙ ФОРУМ ПО ИНФОРМАЦИИ

2012

Том 37

№ 2

---

Выходит ежеквартально

Основан в 1975 г.

## СОДЕРЖАНИЕ

|                                                                                                                                                  |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <b>Бакленд М.</b> Какого рода наукой может быть информатика?                                                                                     | 3  |
| <b>Ма Л.</b> Значения информации: гипотезы и научные последствия<br>трех фундаментальных теорий библиотековедения<br>и информатики               | 10 |
| <b>Шоттон Д.</b> Пять звезд электронных журнальных статей: подход<br>к оценке статей                                                             | 18 |
| <b>Тейссен Р.Я.В., Валтман Л., Эк Н. Я. ван.</b> Научное сотрудничество<br>и расширение научной сети: измерение процессов<br>глобализации в мире | 31 |

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ

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

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

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

**А. Джилкрайст** (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

**Н. Дюсулье** (Франция) — 192 Chemin du Cannet, 06220 Vallauris, France

**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. Фудзивара** (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

Д-р **А.И. Чёрный** (Российская Федерация) — ВИНТИ РАН, 125190, Москва, ул. Усневича, 20. Телекс 411249

# Какого рода наукой может быть информатика?\*

**Майкл БАКЛЕНД**  
(Michael BUCKLAND)

Школа информации Калифорнийского университета, г. Беркли, США

*В течение XX в. было сильное желание развить информатику из библиотекovedения, библиографии и документации, и в 1968 г. Американский институт по документации изменил свое название на Американское общество по информатике. Однако в начале XXI в. отделы (библиотековедения и) информатики вместо этого обратились к общественным наукам. Эти программы рассматривали множество тем, но они были мало успешными в обеспечении согласованного объяснения характера и тематики этой области. Прогресс должен идти в направлении согласованного, унифицированного представления о роли архивов, библиотек, музеев, электронных информационных служб и подобных организаций, если они трактуются как службы по обеспечению информации. Тем не менее, такой подход кажется весьма неполным в обычном понимании обеспечения информации. Вместо вопроса, что такое информатика или чем мы желаем ей стать, мы задаем вопрос, какого рода областью она может быть при определенных наших предположениях о ней. Мы подходим к этому вопросу через объяснение некоторых ключевых слов: «наука», «информация», «знание» и «междисциплинарный». Мы делаем вывод, если информатика имеет отношение к тому, что люди знают, тогда она есть форма культурного знания, и самое большее, наука искусственного.*

## ВВЕДЕНИЕ

В течение XX в. было сильное желание, чтобы обеспечение информационных услуг стало научным, двигалось от библиотековедения, библиографии и документации к информатике. Соответственно, в 1968 г. Американский институт по документации изменил свое название на Американское общество по информатике. Однако в начале XXI в. отделы (библиотековедения и) информатики повернулись вместо этого к общественным наукам. Ведущие программы увеличивали свои размеры и видимость, с умелой гласностью свободно используя слова «информация», «общество» и «технология». «Информационная школа» в настоящее время является именем или уменьшительным именем выбора. Эти программы рассматривают множество важных тем, но они были мало успешными в обеспечении согласованного объяснения характера и тематики этой области. Мудростью организаций считается быть ожидающими новых возможностей, но быть оппортунистичными без согласованного важного рассуждения является неблагоприятным.

Связанная проблема касается анализа информационных служб. Некоторый прогресс может быть достигнут в направлении согласованного, унифицированного взгляда на роли архивов, библиотек, музеев, электронных информационных служб и подобных организаций, если они понимаются как службы, обеспечивающие информацию [1], но такой подход кажется весьма неполным для обычного понимания обеспечения информации. Например, публичные библиотеки делают больше, чем просто обеспечивают информацию. Здесь опять необходимо более глубокое или широкое, или другое объяснение.

## ПОДХОД

Наш подход состоит в том, чтобы рассмотреть некоторые ключевые слова: *наука*, *информация*, *знание* и *междисциплинарный*, и провести различие между словами *научный* (*scientific*), *учёный* (*scholarly*) и *критический* (*critical*).

Хотя слово «наука» иногда используется широко для любого корпуса знаний (например, домашняя наука - домоводство, библиотечная наука - библиотековедение), здесь мы используем его в обычном смысле, обозначающем формальные и физические науки (химия, математика и физика). Наука — это конструктивная инициатива. Как таковое слово *научный* (*scientific*) включает построение модели. Гипотезы и теории разрабатываются, чтобы объяснить и предска-

\* Перевод Buckland M. What kind of science can information science be? // Journal of the American Society for Information Science and Technology. — 2012. — Vol. 63, No. 1. — P. 1-7. — <http://onlinelibrary.wiley.com / DOI 10.1002/asi.21656>.

зять наблюдаемое явление. Слово *учёный* (*scholarly*) включает больше, чем знающий. Оно требует утвердительного поиска для доказательства против чьих-то теорий. Это является истиной для всех областей: в гуманитарных и общественных науках, естественных науках и профессиональных практиках. В этом контексте слово *критический* (*critical*) не подразумевает чего-то враждебного или отрицательного, а поднимает вопросы о важных предположениях и методологических выборах. Насколько выводы были определены или, по крайней мере, подвержены влиянию со стороны определенных предположений или выбора метода? Идеальным является быть научным, учёным и критическим. Чем больше мы сможем приблизиться к этому идеальному, тем более надежными будут наши идеи.

Также важно помнить о различии между вещами и их названиями: описание некоего явления составляет отдельный вопрос, отличающийся от решения о том, как его назвать. Названия могут быть многочисленными, многозначными и нестабильными. Прошлая дискуссия об информации и связанных терминах потерпела неудачу в том, чтобы признать это скорее очевидным различием, и любое декларативное утверждение в форме «Информация есть ...» должно рассматриваться с подозрением из-за отсутствия некоторого объяснения того, к чему оно относится. Другой полезный руководящий принцип – бритва Оккама, принцип, что, при других равных вещах, простейшее объяснение должно быть вообще предпочтительным. Эти принципы обеспечивают основу не только для изучения отдельных понятий информации, но также для рассмотрения того, какого рода областью может быть информатика, и, следовательно, определения вероятного места для программ информационных школ.

## ИНФОРМАЦИЯ

Слово «информация» используется так много, что оно становится доминирующим дискурсом [2]. Сетевой сайт одной информационной школы недавно поместил два поразительных заявления: «161 эксабайт новой информации создается ежегодно» (они означают цифровые биты) и «Информация: энергия для преобразования мира» (они не означают цифровые биты). Неопределенность и несовместимость являются достоинствами лозунгов, а использование «слов-хамелеонов», которые придают различные цвета разным контекстам, позволяет читателям гибко понимать, чего они хотят. Однако, когда нужна ясность, требуются более тщательные определения.

Наше первое ограничение – это ограничить наше использование информации до ее традиционной связи с познанием и изучением человека. Это отличает нашу тематику от других важных областей, также использующих название «информатика». Одна из них – вычислительная техника, касающаяся теории и применения алгоритмов. Другая область, касающаяся энтропии, вероятности, теории информации Шеннона - Уивера, физических моделей (*in-form-ing*) и связанных тем, иногда относится к «физике информации». Также слово «информация», конечно, используется в информационной технологии, информационных и коммуникационных технологиях, но оно больше ограничено практикой использования электроники для коммуникации и вычисления. Эти другие сферы здесь не рассматриваются. Вместо этого мы имеем дело с теми сферами, вообще понимаемыми как тематика библиотекведения и ин-

форматики и интересы Американского общества по информатике и технологии. Более широкий и подробный анализ многочисленных областей с некоторым интересом к информации см. в [3].

Джонатан Фурнер [4] мудро напоминает нам, что для каждого из множественных значений слова «информация» всегда имеется другое удовлетворительное, более специальное слово. Информационные исследования не требуют использования слова «информация»! Другое движение состоит в сортировании разнообразных использований слова «информация» по категориям, включающим:

*Информация как знание* для сообщенного знания, которое было изучено в результате информирования.

*Информация как процесс* для становления информированным, для изучения.

*Информация как вещь* для битов, байтов, книг, звукозаписей, изображений и любых физически ощущаемых вещей в качестве значимых. Слово «документ», которое не было исторически ограниченным текстовыми средствами, может использоваться в качестве технического термина для «информации как вещь» [1, 5, 6].

Начиная с этой последней категории, информация как вещь, мы можем спросить, что документы делают или, правильнее, что люди делают с информацией в качестве вещи, с документами, т.е. так сказать, с данными, записями, текстами и средствами каждого вида.

## ИСПОЛЬЗОВАНИЕ ДОКУМЕНТОВ

Если мы посмотрим, то обнаружим, что документы широко применяются для разных целей. Правительства применяют документы для контроля за нами, требуя использования паспортов, уплаты подоходного налога, лицензии водителей и т.п. Школы применяют учебники и учебные планы, чтобы управлять студентами и преподавателями. Религии используют священные тексты для постепенного внушения веры и влияния на поведение. Торговцы вкладывают большие инвестиции в рекламу, чтобы влиять на то, чтобы мы покупали. Политики используют лозунги и политические заявления, чтобы завоевать голоса и привлечь финансовую поддержку и поддержку электората. Эстрадные артисты используют разные средства, чтобы развлекать нас, и обычно, чтобы получить плату от нас. Отдельные лица применяют сообщения, чтобы общаться, а общественные средства, чтобы привлечь внимание. Музеи дают объяснения нашего наследия через селективное представление и искусную интерпретацию артефактов. Библиотеки обеспечивают доступ к фондам документов и т.д. Любой может сделать такой перечень, и этот перечень быстро становится длинным.

Рассмотрение этого или любого подобного перечня напоминает нам о некоторых других пунктах:

1. Документы проникают в общество и формируют нашу жизнь. Зависимость от документов растет со временем. Современные экономики основываются на все растущем разделении труда и на существовании рынков, которые зависят от коммуникации и документации, и, в свою очередь, успешно облегчаются техническими инновациями (письмо, печать, телеграф, радио, Интернет и т.д.). Как говорил Патрик Уилсон, мы всё больше зависим от «знаний из вторых рук» [7].

2. Использование информации и информационное поведение обычно понимаются как связанные с ин-

дивидуумом, которому должно нравиться быть информированным. Однако, как явствует из перечня, это только небольшая часть истории. Много случаев использования документов инициируется не пользователем, а широким и разнообразным множеством очень активных агентов (правительств, школ, религий, торговцев и т.п.) в целях различения, а иногда конкуренции.

3. Самая общая форма поведения, связанного с информацией, – это просто привлекательные вещи, минимально активная роль. Оно (поведение) может быть непреднамеренным (как, когда мы слышим гром), неожиданным или неосознанным (когда является подсознательным).

4. Использование документов может включать – но не делать меньшим – нахождение фактов, поиск информации или решение проблемы. Как явствует из этого перечня, программы (повестки) и значения различаются. Публичные библиотеки являются не просто информационными службами, по крайней мере, в любом простом или обычном смысле. «Одна из этих вещей, которую публичные библиотеки должны делать честно, – это понимать, что их миссией, их работой является создание сообщества», – утверждает Мартин Гомес [8].

Если мы посмотрим на этот приведенный выше перечень или на любой подобный перечень, разумно спросить, какой термин может охватить этот круг информационной деятельности. Общей чертой можно считать то, что они являются культурными видами деятельности. Здесь мы используем «культуру» не в популярном смысле высокой культуры, а в широком научном смысле, применяемом в антропологии. Классическим является определение, данное Тейлором в 1871 г.: «Культура, или цивилизация, взятая в её широком этнографическом смысле, есть комплексное целое, включающее знание, веру, искусство, мораль, закон, обычаи и любые другие способности и привычки, приобретенные человеком как членом общества» [9, с.1]. Простейшее предположение затем состоит в том, что использование информации, когда мы говорим об информации как о вещи, следует рассматривать как активное занятие в культурной сфере.

## ЗНАНИЕ

Теория знания была доминирующей в аналитической психологии с акцентом на истину утвердительных высказываний и знаний как справедливой истинной веры [10]. Этот подход проблематичен по нескольким причинам. Мы можем усомниться относительно прилагательного «справедливая», так как никто, вероятно, не считает, что он исповедует несправедливую веру. Критерий истины также не подлежит проверке. В обычной речи «истина» стремится подразумевать совместимость с некоторой объективной реальностью, а субъективное знание объективной реальности с философской точки зрения подозрительно, и на практике истина сокращается до совпадения с некоторой другой прежней верой или предположением.

Утвердительное знание (справедливая истинная вера) иллюстрируется в следующей выдержке из «Стандфордской философской энциклопедии» [11]:

«Предположим, например, что Джеймс, отдыхающий на скамейке в парке, наблюдает за собакой, которая находится от него почти в 8 ярдах и грызет кость. Он думает так

5. Вон там собака.

Предположим далее, что та, которую он принимает за собаку, на самом деле собака-робот, настолько совершенная, что только посредством зрения нельзя отличить ее от настоящей собаки ... Эти данные предположения, естественно, ошибочные. Но предположим далее, что как раз в нескольких шагах от собаки-робота находится настоящая собака. Сидящая за кустом, она скрыта от взора Джеймса. Дальнейшее предположение – вера Джеймса истинна. Итак, еще раз, то, что мы имеем перед нами, является справедливой истинной верой, которая ... дает нам ошибочный результат о том, что знает Джеймс (5)».

Аналитическая философия такого рода мало соответствует повседневным реальностям общества, насыщенного документами, нашей неизбежной зависимости от знания из вторых рук и вечной потребности решать, кто и во что верит. Известное пособие по логике XVII в. обобщило ситуацию довольно хорошо:

«... широкое различие должно делаться между двумя видами истин: одной, которая связана просто с характером вещей и их неизменной сущностью независимо от их существования; другой, которая связана с существующими вещами и особенно с человеческими случаями и событиями, ...

В первом виде истин, поскольку всё является необходимым, ничто не является истиной, которая не является истиной вообще; и таким образом, мы можем сделать вывод, что вещь является ложной, если она ложна в единственном случае. Но если мы думаем о следовании тем же правилам в вере относительно человеческих событий, мы будем всегда, за исключением случайности, судить ложно и делать тысячи ложных умозаключений о них. Для этих событий как случайных по своей природе должно быть нелепым искать в них необходимую истину ...» [12, с. 345-346].

В 1946 г. Джилберт Райл писал: «Философы не справедливы к отличию, которое вполне известно всем нам, между знанием, что что-то является случаем, и знанием, как делать вещи [13, с. 4]. Он утверждает, что *знание*, как не может быть определено в терминах *знания*, *что*, и что *знание*, как было логически прежде *знания*, *что*. Но этого недостаточно. Теорию знания необходимо расширять далее до другого отличия, вполне известного всем нам: *знание о*. В нашей повседневной жизни мы оперируем с неизбежно несовершенным, неполным и неопределенным знанием. Мы должны постоянно принимать решения в зависимости от этого или того документа. В реальной жизни мы имеем несовершенное *знание о* и должны полагаться больше на веру, чем на истину. Учитывая это положение, важность расположения наиболее подходящих доступных документов для нас самих или для других, существенно касающихся библиотекосведения и информатики, является очевидной. В этой ситуации различие между знанием и верой оказывается сомнительным, а предполагаемое знание, прежде занятое истиной единичных сентенций, становится неправдоподобной теоретической основой.

## СТАНОВЛЕНИЕ ИНФОРМИРОВАННЫМ

Оставшаяся категория *информация как процесс* связана с сообщением знания, с обучением. До тех пор, пока мы касаемся понимания, а не просто запоминания, обучение зависит от того, что мы уже знаем. Обучение является прирастающим, изменение в том, что мы знаем, а не просто в добавлении, за исключением, кажется, в исследовании по библиотекосведению и информатике, где

мы обнаруживаем основной дефицит. Подробный анализ содержания литературы по библиотековедению и информатике, проведенный Алланом Конрадом, показал, что только 5,6% выборки из 413 изученных канонических текстов соответствовало принципу, что обучение является прирастающим; большая часть текстов (88,8%, включая 83% в подмножестве, категоризированном как когнитивные исследования) либо игнорировала этот принцип, либо только символически упоминала его; 5,6% определенно или безоговорочно опровергало его [14, с. 499-569, особенно с. 508].

При таком стремлении этой области к общественным наукам и определенном разговоре о «когнитивном повороте» [15] эти результаты являются поразительными. Мы можем размышлять о причинах. Одно соображение состоит в том, что очень трудно на практике учесть то, что отдельные лица уже знают. Другой фактор состоит в том, что формальные и алгоритмические методы в силу своей природы сопротивляются включению культуры [16]. Ньютоновская физика не принимает во внимание небеса или ад, а информатика в основном фокусируется на системах хранения и поиска информации, фактически, скорее, на системах обеспечения документов, чем на системах, которые информируют [1]. Третья причина, связанная с некоторыми достойными похвалы исключениями (такими, как у Аллена Брайса и Кэрола Култау), так называемый когнитивный поворот набирает тенденцию вместо того, чтобы быть, скорее, узко основанным на когнитивной науке поворотом. (Уитнесс часто ссылается на статьи по искусственному интеллекту, а не на широкие области образовательной психологии.) Частая ссылка на «состояние знания» подразумевает сомнительное упрощение, поскольку каждый раз мы вспоминаем что-то, что мы создаем слегка отличающимся воспоминанием. Сёрен Брайер [17] характеризует эту ситуацию как следующую. На современную доминирующую парадигму сильно влияет когнитивная наука, которая является логической и алгоритмической исследовательской программой, изучающей обработку информации в людях, животных и машинах. Этот подход основан на кибернетике Винера, информационной теории Шеннона – Уивера, логике, теории множеств и вычислениях. Он является неадекватным, поскольку не умеет приспособливаться к культурным реальностям знания и его передачи, феноменологической сложности восприятия и понимания или взаимодействию общественного и персонального. В результате возникает общая путаница среди многих альтернативных значений слова «информация» и подходом к информационному поведению, который непривлекателен для коммуникации и обучения.

## ЯЗЫК И ФАКТЫ

Информационный поиск, широко рассматриваемый (наряду с библиометрией) как наиболее научная часть информационных исследований, сильно зависит от алгоритмических операций на тексте, особенно совместной встречаемости специальных слов (фактически, цепочек знаков) в запросе и искомым документах. Эти методы очень полезны, несмотря на некоторые недостатки, возникающие из-за слов, имеющих множественные значения и различные формы, из-за разных слов, имеющих одно и то же написание, и из-за нестабильных значений. Наоборот, человеческая коммуникация зависит от культурных кодексов и значения. Роберт Фэрторн [18] имел правильные взгляды на эти вопросы с

его тщательным различением между *упоминанием* и *значением* и его объяснением непреодолимой устарелости предметного индексирования. Язык развивается в диалоге и речи. Индексатор обязательно должен оглядываться назад, поскольку индексационные термины должны быть основаны на употреблении, уже установленном в прошлой речи. Но индексатору также следует быть вперед смотрящим, поскольку индексирование предназначено для дальнейшего применения. Значения слов продолжают развиваться во времени, но индексационный термин, приписанный в какой-то фиксированный момент времени, уходит в прошлое, тогда как речь, язык и индексатор идут вперед [19, 20].

Когда культура и язык не поддаются алгоритмам и формальным методам, больший прогресс может быть достигнут, если мы сможем свести буквальное к фактическому. Так думал Поль Отле. Он считал, что книги и статьи становятся неэффективными, чрезмерно самоуверенными и дублируемыми. Его идеей было извлекать факты из текстов, подобно гороху из стручков, и организовывать факты в авторитетную семантическую сеть, используя сжатые единичные фактические предложения («монографы»), описанные, расположенные и совместно связанные с помощью Универсальной десятичной классификации (УДК) [21]. Он говорит, что результат может быть разделенным как коллективное расширение разума. (Понимание самим Отле «мирового разума», разделяемое Вильгельмом Оствальдом и Х. Уеллсом, представлялось общественным ресурсом, больше похожим на строго изданную Википедию, чем на самостоятельную вещь, подобную компьютеру "Hal" в 2001 г.: A Space Odyssey.)

Однако в то же самое время, когда Отле обобщил свои идеи в трактате "Traité de documentation", опубликованном в Брюсселе в 1934 г. [22], Людовик Флек в Польше выразил очень отличающийся взгляд в своем труде «Генезис и развитие научного факта», опубликованном в следующем 1935 г. [23]. Флек утверждает, что факты, найденные в популярных энциклопедиях, были сверхпростыми, когда сокращались до простых предложений из контекста и изолировались от объяснительных повествований. Далее, факты возникают только в триадической связи понятия (индивидуального) и преобладании культурного типа мышления (*Denkkollektiv*), которые оба являются как годными, так и ограниченными. Как говорил Флек, даже научные факты, долго противоречащие парадигмам и научным революциям Томаса Куна и подтверждающие археологию знания Мишеля Фуко, являются культурно подходящими конструкциями. Даже небольшое внимание к интеллектуальной истории иллюстрирует его случай. Парацельс, врач эпохи Возрождения, который боролся за современную науку, первым используя в медицине химические вещества и проявляя пристальное внимание к размеру дозы, был так погружен в средневековую алхимию, что испытывал недостаток адекватных понятий и терминологии [24]. Он бы не понял наших современных медицинских текстов, а мы бы не поняли его.

## ПОДХОДЯЩЕЕ РАСПОЛОЖЕНИЕ ДОКУМЕНТОВ

Ответ В. Соуминен [25] на вопрос «Что представляет собой хороший библиотекарь?» - был, что хороший библиотекарь – тот, который достигает подходящего расположения документов для читателя. Существует

много различных читателей, каждый из которых имеет множество интересов, для которых существует много документов, и что читатели, интересы и документы - все являются очень нестабильной, довольно сложной задачей, но понятие является, в принципе, привлекательным. Конечно, библиотечное дело и информатика во многом касаются подходящего расположения документов тем или иным способом.

Библиография и библиографическое описание связаны с установлением подходящих расположений двумя способами: отношения между документами устанавливаются через описания, дескрипторные перечни и индексы, и эти описания, дескрипторные перечни и индексы используются для определения соответствующих документальных значений с некоторой целью [26].

Однако сложные информационно-поисковые системы имеют такие же важные свойства, как и библиография. Весь механизм выбора (для поиска и фильтрации) включает цепочки только двух примитивных типов операции: модификация документов (включая развитие индексов) и их (пере)расположение (сортировка, ранжирование, кластеризация и т.п.) [27, 28]. Информационный поиск является алгоритмическим, количественным и весьма полезным, но является ли он научным? [29]. Операции системы выбора зависят от теории множеств и релевантности. Теория множеств представляется удобным упрощением, поскольку документы не являются в действительности по отдельности различными по содержанию или значению. Большая проблема состоит в том, что нет такой вещи, как релевантность, по крайней мере, ничего осязаемого, когда становится яснее, если мы заменяем слово «подходящий» на «релевантный», что мы можем сделать без изменения смысла. «Релевантность» может казаться более научным словом, поскольку оно имеет формальное значение необходимости в логике и поскольку 50 лет релевантностью измеряют оценку информационного поиска. На практике документы упорядочиваются с использованием некоторого заменителя релевантности. Эти заменители ранжируются от совместной встречаемости цепочек знаков до использования оценок о релевантности третьей стороной, иногда изменяются после начального поиска из-за субъективного восприятия тех, для которых поиск выполняется. Вообще, несмотря на то, является ли позиция отступления темой (или употреблением слова), она остается подобной [30]. Немного удивительно, что определения и литература по релевантности остаются упорно проблематичными в течение 50 лет, несмотря на поддерживающие усилия таких очень талантливых и мотивированных учёных. Естественная наука (например, химия или физика) требует измеряемого физического свойства. Формальная наука (например, логика или математика) должна требовать ясного и строгого определения. Для более поддающихся влиянию общественных наук характерно то, что ни то, ни другое для них недоступно, и необходимо делать всё возможное, по крайней мере, с неудовлетворительными заменителями.

Библиометрия, главным образом базирующаяся на анализе ссылок, является другим эпицентром количественного определения в информатике. Здесь и мотивы, и значение отдельных актов ссылок, стремятся оставаться неясными, за исключением неопределённых общих терминов, так что библиометрия имеет сходство с информационным поиском, в котором виртуозное вычисление не основывается на твердых основах. Библиометрия и информационный поиск несут методы,

разработанные в и для формальных (логических, хорошо определенных) сред и используют их на объектах и в средах, которые не являются формальными, логическими или хорошо определенными. Это приводит к полезным результатам, а также обязательно к компромиссным, несовместимым процессам.

## МЕЖДИСЦИПЛИНАРНЫЙ ХАРАКТЕР

Иметь междисциплинарный характер широко рассматривается, как быть хорошей вещью, а иногда таковою и являться. Хорошим практическим примером было бы, когда желаемая академическая программа не утверждается, представление альтернативной программы, созданной в качестве междисциплинарной программы, может быть успешным. Тем не менее, слова, начинающиеся с «интер» и «меж», вообще обозначают позицию слабости (например, интервал, интермиссия, междоусобица и промежуток) и указывают на что-то расположенное между другими, более существенными предметами.

Персональный взгляд состоит в том, что требования (заявления) в университетской среде быть междисциплинарной имеют привлекательность среди планирующих органов, но что во время экономического кризиса политическая власть старается находиться в хорошо сложившихся дисциплинах. Таким образом, аргументируя требование к ресурсам, основанное на том, быть междисциплинарной или быть новой дисциплиной, - означает, вообще-то, выбор для занятия слабой позиции.

К счастью, для информационных исследований существует строгая альтернатива — общественная потребность. Кто хочет иметь дело с несведущим механиком, врачом с устаревшими медицинскими знаниями, плохо информированным менеджером, выпешшими из употребления пособиями или скрытым правительством? За некоторыми исключениями, заметно связанными с тайной и секретностью, мы все имеем существенный «жилетный» интерес в интеллектуальном обществе. Нам нужны люди, которые хорошо информированы, которые знают о том, что они делают. Главные общественные потребности обычно комплексные. Кто бы ни принялся решать эти потребности, он должен быть методологически многосторонним в способе, который является недостаточно охваченным «междисциплинарной» наукой. Ирония заключается в том, что большинство уважаемых научных отделений (например, история, химия и языки), возникших в XIX в., осознавали общественные потребности как национально-государственные.

Книга Брайера «Кибернетика: почему информация является недостаточной!» [17] представляет собой необычно эрудированную, полную и связанную теоретизацию природы междисциплинарной. Является ли эта книга междисциплинарной? Она распространяется широко на несколько областей, включая биологию, кибернетику, психологию, семиотику и другие области, так что это очевидно. Но сказать так, значит упустить более важную суть, что она является связанной унифицированной теорией для существующей области.

Каждая научная специальность развивает свою собственную культуру знания, языка, ценностей и социальных структур. В результате они обязательно более или менее отличаются одна от другой по тематике и потенциально несовместимы или, по крайней

мере, противоречивы. Ни одна специальность, вероятно, не должна отдавать предпочтение единообразной культуре (эпистемология, терминология) для развития ее собственной исконной культуры, так что должно ожидать напряжение между желанием выгоды (пользы) от совместимости с другими специальностями и дискомфортом от связи с более или менее чуждыми культурами других специальностей. Линдхолм-Романчук, последний, кто изучал поток идей в/среди дисциплин, имел мнение, что самая продуктивная позиция – быть основательно подготовленным в своей собственной области, а затем продвигаться с исследованием к границам или через границы с другими областями (личное сообщение, 1994 г.).

## ЗАКЛЮЧЕНИЕ

«Информация» и другие неопределенные и/или многозначные слова могут быть очень существенными в лозунгах и риторике.

Информатика была использована для обозначения различных областей, которые мы можем различать с использованием разных названий: библиотекведение и информатика, вычислительная техника, физика информации, энтропия и т. д., – а информационная технология, означающая электронную технологию, применялась для коммуникации и вычисления. Только первая из них непосредственно связана со знанием и обучением.

Стать людям лучше информированными (обученными, более знающими) есть, или должно быть, центральной заботой информационных исследований, а информационные службы на практике больше непосредственно связаны со *знанием о*, чем со *знанием, как* или *знанием, что*. Знание в повседневной жизни есть вера, есть культура, и не обязательно хорошо объясненная или истинная в каком-то строгом смысле. Одно последствие состоит в том, что тонкости аналитической философии обеспечивают неподходящую основу для теоретизации информатики.

В повседневной жизни мы сильно, и всё больше и больше, зависим от знания из вторых рук. Сами мы мало что можем определить из того, что нам необходимо знать, из первых рук, непосредственно из опыта. Мы вынуждены зависеть от других, в основном через документы. Соответственно, существует множество агентств, стремящихся влиять на нашу жизнь и использующих документы как свои средства для достижения своих меняющихся, а иногда спорных целей. В этом потоке информации мы должны выбрать и решить, чему доверять. То, что мы думаем о документе, влияет на наше использование его, и более важно, что наше использование документов влияет на то, что мы думаем. Сюзанна Брие осознавала эти проблемы, когда она писала в 1951 г. о документации не только как о «настоятельной потребности нашего времени», но и как о «новом *культурном* методе» [31, дополнительно подтверждено в 32].

Информационный поиск и библиометрия, оба как очень полезные методы, являются количественно определяемыми и техническими, но не научными в обычном смысле, поскольку они базируются на слабо определенных основах. Если информатика – это наука, то она является искусственной [33], а не естественной наукой (подобно физике) или формальной наукой (подобно математике). Патрик Уилсон был прав, когда писал, что информационные исследования включают широкий круг общественных наук (и гуманитарных) и некоторых высоко специализированных технических наук [34, 35].

Эти выводы должны раздражать многих, кто в XX в. был решительно настроен создать важную науку информатику. Ответ должен быть таким, что если проблема является важной, то характер проблемы должен определять методологию, а не наоборот. Некоторым утешением должно быть то, что любое (пере)создание области освещает как возможности, так и недостатки. Некоторые из них, которые приходят на ум, – это то, что если методы аналитической психологии в планируемом знании оказываются бесплодными для наших потребностей, тогда упор на знание, как на веру, так и на культуру, должен быть плодотворным. Если мы придерживаемся функциональной точки зрения и замечаем, что взаимосвязи между документами долго изучались в гуманитарных науках, тогда библиография может быть обогащена рассмотрением ее как формы паратекста, и наоборот. Если использование алгоритмов зависит от полезного упрощения, показывается более глубокое исследование последствий этого компромисса. Интернет расширил вопрос решения, каким документам доверять, но изучение того, насколько доверие влияет на использование документа, требует дополнительного изучения того, насколько использование документа влияет на доверие, и если возрастающая природа становления информированным отрицается, должно быть много сделано, чтобы перевести акцент со служб, обеспечивающих информацию, на системы, которые информируют.

По этим причинам, изложенным в настоящей статье, установлено, что информатика касается культурного обязательства. Формальные и количественные подходы весьма важны, но область сама по себе неисправимо культурная. Формальные и количественные методы, какими бы полезными они ни были, никогда не могут быть больше, чем высокоценными вспомогательными ролями. Характеристика информационного поиска и библиометрии как наук искусственного является описанием, а не критикой.

Эти выводы не направлены на другие разного рода информационные исследования, особенно вычислительную технику, физику информации или информационную технологию, которые не связаны непосредственно с тем, чему верят люди.

Наконец, мы можем видеть, что наши аргументы имеют кругообразную форму. Существоют последствия того, что однажды мы решили признать центральное понятие информации как имеющее дело со знанием и обучением. Во-первых, произошло отделение, по существу, от зон свободного знания, занятых вычислительной техникой, физикой информации и информационной технологией. Во-вторых, любое понятие информационных исследований, включая то, что и как мы знаем, может быть культурным вопросом. В-третьих, полезные формальные и количественные средства зависят от значительного упрощения компромиссов, необходимых для снижения субъективных и культурных качеств области. Наконец, принятие культурного контекста информатики должно привести к более реалистичному и более эффективному вкладу в наше насыщенное документами общество.

*Благодарность.* Ранние версии этой статьи были представлены на Конференции документальной академии, проведенной Информационным колледжем Университета Северного Техаса, г. Дентон, 19 марта 2010 г., и в Школе библиотекведения и информатики Университета Южной Каролины, 7 апреля 2011 г.



## ЛИТЕРАТУРА

1. *Buckland M.* Information and information systems. – Westport, CT: Praeger, 1991.
2. *Day R.* The modern invention of information: Discourse, history and power. – Carbondale: Southern Illinois University Press, 2001.
3. *Machlup F., Mansfield U.* (Eds.) Study of information: Interdisciplinary messages. – New York: Wiley, 1983.
4. *Furner J.* Information studies without information // *Library Trends*. – 2004. – Vol. 52. – P. 427-446.
5. *Buckland M.* Information as thing // *Journal of the American Society for Information Science*. – 1991. – Vol. 42. – P. 351-360. – <http://people.ischool.berkeley.edu/~buckland/whatdoc.html>
6. *Buckland M.* What is a 'document'? // *Journal of the American Society for Information Science*. – 1997. – Vol. 48. – P. 804-809. – <http://people.ischool.berkeley.edu/~buckland/whatdoc.html>
7. *Wilson P.* Second-hand knowledge: An inquiry into cognitive authority. – Westport, CT: Greenwood Press, 1983.
8. Institute of museum and library services. The future of museum and libraries: A discussion guide. – Washington, DC: Author, 2009.
9. *Tylor E.B.* Primitive culture. – London: J. Murray, 1871.
10. *Chisholm R.M.* Theory of knowledge (3<sup>rd</sup> ed.). – Englewood Cliffs, NJ: Prentice-Hall, 1989.
11. Stanford Encyclopedia of Philosophy. The analysis of knowledge. – 2006. – <http://plato.stanford.edu/entries/knowledge-analysis/>
12. *Arnauld A.* Logic; or the art of thinking: Being the Port Royal logic (T.R. Baynes, Trans.). – Edinburgh, Scotland: Sutherland and Knox (Оригинальная работа опубликована в 1662 г.).
13. *Ryle G.* Knowing how and knowing that // *Proceedings of the Aristotelian Society*. – 1946. – Vol. 46. – P. 1-16.
14. *Konrad A.* On inquiry: Human concept formation and construction of meaning through library and information science intermediation. (Неопубликованная докторская диссертация). University of California, Berkeley, 2007. -<http://escholarship.org/uc/item/Is76b6hp>
15. *Ingenversen P., Järvelin K.* The turn: Integration of information seeking and retrieval in context. – Dordrecht, The Netherlands: Springer, 2005.
16. *Eklbia H.R.* Artificial dreams: The quest for non-biological intelligence. – New York: Cambridge University Press, 2008.
17. *Brier S.* Cybersemiotics: Why information is not enough! – Toronto: University of Toronto Press, 2008.
18. *Fairthorne R.A.* Temporal structures in bibliographic classification / J.A. Wojciechowski (Ed.) // *Conceptual basis of the classification of knowledge: Proceedings of the Ottawa Conference on the Basis of the Classification of Knowledge*. – Pullach, Germany: Verlag Dokumentation, 1974. – P. 404-412.
19. *Buckland M.* Naming in the library: Marks, meaning and machines / C.Todenhagen, W. Thiele (Eds.) // *Nominalization, nomination and naming in texts*. – Tübingen, Germany: Stauffenburg, 2007. – P. 249-260. – <http://heople.ischool.berkeley.edu/~buckland/naminglib.pdf>
20. *Buckland M.* Obsolescence in subject description // *Journal of Documentation* (в печати).
21. *Frohmann B.* The role of facts in Paul Otlet's modernist project of documentation / W.B. Rayward (Ed.) // *European modernism and the information society*. – Aldershot, England: Ashgate, 2008. – P. 75-88.
22. *Otlet P.* *Traité de documentation*. – Brussels: Editiones Mundaneum, 1934.
23. *Fleck L.* Genesis and development of a scientific fact. – Chicago: Chicago University Press, 1979. (Оригинальная работа опубликована в 1935 г.).
24. *Ball P.* The devil's doctor: Paracelsus and the world of Renaissance magic and science. – New York: Farrar, Straus and Giroux, 2006.
25. *Suominen V.* Filling empty space: A treatise on semiotic structures in information retrieval, in documentation, and in related research. – Oulu, Finland: Oulu University Press, 1997.
26. *Wilson P.* Two kinds of power: An essay on bibliographical control. – Berkeley: University California Press, 1968.
27. *Buckland M., Plaunt C.* On the construction of selection systems // *Library Hi Tech*. – 1994. – Vol. 48. – P. 15-28.
28. *Plaunt C.J.* A functional model of information retrieval systems. – University of California, Berkeley, 1997. (Неопубликованная докторская диссертация).
29. *Neill S.D.* Dilemmas in the study of information: Exploring the boundaries of information science. – Westport, CT: Greenwood Press, 1992.
30. *Buckland M.* Relatedness, relevance, and responsiveness in retrieval systems // *Information Processing and Management*. – 1983. – Vol. 19. – P. 237-241.
31. *Briet S.* What is documentation? – Lanham, MD: Scarecrow Press, 2006. (Оригинальная работа опубликована в 1951 г.).
32. *Day R.* 'A necessity of our time': Documentation as 'cultural technique' / S. Briet (Ed.) // *What is documentation?* – Metuchen, NJ: Scarecrow Press, 2006.
33. *Simon H.A.* The science of the artificial. – Cambridge, MA: MIT Press, 1996.
34. *Wilson P.* Bibliographical R&D / F. Machlup, U. Mansfield (Eds.) // *The study of information: Interdisciplinary voices*. – New York: Wiley, 1983. – P. 389-397.
35. *Wilson P.* The future of research in our field / J.E. Olaisen, J.E. Munch-Petersen, P. Wilson (Eds.) // *Information science: From the development of the discipline to social interaction*. – Oslo, Norway: Scandinavian University Press, 1996. – P. 319-323.

# Значения информации: гипотезы и научные последствия трех фундаментальных теорий библиотековедения и информатики\*

Лай МА  
(Lai MA)

Школа библиотековедения и информатики,  
Университет шт. Индиана,  
г. Баумингтон, США

*Исследуется вопрос, «что такое информация», путем сравнения значения термина «информация» и эпистемологических гипотез трех теорий в библиотековедении и информатике: «модель Шеннона-Уивера», интерпретация Бруксом Мира 3 Поппера и модель «данные-информация-знание-мудрость» (ДИЗМ). Показывается, что термин «информация» в этих теориях относится к эмпирическим сущностям или событиям и понимается как причинно влияющий на мнения людей. Утверждается, что эпистемологические гипотезы приводят к небрежности в культурных и социальных аспектах построения информации (т.е. то, как что-то считается или нет информацией) и бесспорной природы науки в исследовательских методологиях.*

## ВВЕДЕНИЕ

Обсуждение термина «информация» не является новым в библиотековедении и информатике. Назовем несколько работ, опубликованных в 2000-х гг.: Капурро и Йорланд [1] представляют исчерпывающий обзор понятия информации, включая его использование в естественных, гуманитарных, социальных науках и информатике; Дей [2] выполняет критический анализ информационного дискурса в европейской документации, кибернетике в век «виртуальности»; Фроманн [3] пытается выявлять информацию; Бремен [4] концептуализирует информацию как строительную силу общества для обсуждения информационной политики; Флориди [5] обсуждает рассеяние информации в каждом аспекте человеческой жизни; Сааб и Рисс [6] понимают информацию как процесс ощущения и осмысления; Марчионини [7] погружает информацию в киберпространство; Глесик [8] прослеживает информацию от барабанного боя до квантового вычисления; и, конечно, Бейтс [9-11] и Йорланд [12,13] полемируют о природе информации.

Почему важно задавать вопросы об информации? Потому что значения термина «информация» варьируются в различных ситуациях: информация может быть новостями о выверке стоимости облигаций правительства США или лекцией в феноменологии Хайдеггера, или краткой беседой о собаке вашего соседа. Значение термина «информация» кажется неоднозначным, поскольку он обычно используется для представления многообразия вещей, событий и выражений, а иногда

знания об этих вещах, событиях или выражениях. Когда мы используем термин «информация» в нашей повседневной речи, мы обычно не задумываемся или не поясняем, что он означает; конечно, мы обычно понимаем значение термина «информация» без долгих раздумий. Этот термин подобным образом представляется в некоторых научных исследованиях и концептуализациях. Значения термина «информация» и их гипотез варьируются в различных текстах и ситуациях. Значит, важно задавать вопросы об информации, так как неоднозначность значений этого термина подразумевает, что: а) термин «информация» используется без пояснения того, к чему он в действительности относится или что он означает, б) что «информация» означает в некоторых контекстах, может быть неправильно понято по коннотациям (сопутствующие значения) и в) использование термина «информация» может быть идеологическим в том смысле, что оно может маскировать или искажать «действительное значение» своими коннотациями.

Эта статья подходит к вопросу, «что такое информация», путем изучения онтологических и эпистемологических гипотез понятий информации в трех фундаментальных теориях библиотековедения и информатики: модель Шеннона-Уивера [14], интерпретация Бруксом [15] Мира 3 Поппера и модель «данные-информация-знание-мудрость» (ДИЗМ) [16,17]. Данная статья объединяет эти три теории для сравнения различных значений термина «информация» между собой и предположения об их сходстве с точки зрения их онтологических и эпистемологических гипотез и убеждений. Утверждается, что эта проблема не является вопросом многозначности значений или концепций информации, а скорее всего тем, что это разнообразие на самом деле дает неверное представление об их общих онтологических и эпистемологических гипотезах и убеждениях, которые могут привести к небрежности в компонен-

\* Перевод Ma L. Meanings of information: The assumptions and research consequences of three foundational LIS theories// Journal of the American Society for Information Science and Technology. — 2012.—Vol. 63, No. 4.— P. 716-723.— <http://online-library.wiley.com/doi/10.1002/asi.21711/pdf>

тах информации (т.е. как что-то считается информацией или нет) и бесспорной природы науки в исследовательских методологиях.

### Модель Шеннона-Уивера

Статья Клода Шеннона «Математическая теория коммуникации», как правило, признается фундаментальным текстом в информационной теории. Понятие информации и соответствующая единица измерения «бит» (бинарное число) оказали значительное влияние на сферу квантовой теории в теоретической физике и на развитие цифровых и квантовых вычислений, например предположение Джона Уилера «она из битов» [8, 18, 19]. Большая часть толкований информационной теории в библиотековедении и информатике, тем не менее, основана на объяснении Уорреном Уивером теории Шеннона. В 1983 г. Шоу и Дэвис уже утверждали, что «большая теоретическая работа в информатике основана на модели коммуникации Шеннона-Уивера» [20, с. 71]. Позже Колбах [21, с. 48] утверждает, что Шеннон и Уивер являются «отцами современной теории информации и коммуникации» и что понятие неопределенности в их работе «соответствует многим аспектам нашей жизни». Заметим, хотя обе статьи не относятся к информационной теории, а скорее «модель Шеннона-Уивера» является показателем того, что толкования понятий в информационной теории основаны на эссе Уивера. Следовательно, нам необходимо понять, как Уивер развивает теорию Шеннона в общую теорию коммуникации, такую, которую мы можем понимать через понятие информации, которое затем сообщает об использовании этих текстов библиотековедением и информатикой.

«Математическая теория коммуникации» Шеннона и Уивера по сути не является соавторской книгой. Эта книга состоит из двух эссе: «Последние вклады в математическую теорию коммуникации» Уивера и более ранняя статья Шеннона с новым названием «Математическая теория коммуникации». Эссе Шеннона было перепечатано из журнала *Bell System Technical Journal* (июль и октябрь 1948 г.), а краткое эссе Уивера было первоначально опубликовано в *Scientific American* (июль 1949 г.) [14]. Темы обоих эссе обслуживают разные аудитории: Шеннон пишет о технической инженерной проблеме передачи сообщений (например, как цепочка символов может успешно передаваться от одной машины к другой), тогда как Уивер пытается предложить *общую* теорию коммуникации, основанную на статье Шеннона.

Система коммуникации Шеннона теоретизирует проблемы передачи сообщений от информационного источника к месту назначения через канал. Технической про-

блемой является выбор лучших сообщений (т.е. символического представления для кодирования) и минимизация шума, который может помешать процессу передачи, например во время телеграфной передачи от одного терминала к другому. Шеннон утверждает во введении к своему эссе:

«Фундаментальной проблемой коммуникации является проблема воспроизводства в одной точке либо точно, либо приблизительно сообщения, выбранного в другой точке. Часто эти сообщения имеют *значение*, т.е. они относятся или коррелируются в соответствии с некоторой системой с определенными физическими или концептуальными объектами. Эти семантические аспекты коммуникации являются irrelevantными инженерной проблеме» [14, с. 31].

Хотя Шеннон явно утверждает, что семантические аспекты коммуникации являются irrelevantными для технических проблем, Уивер, тем не менее, рассматривает эту техническую модель как основу общей теории коммуникации и как «реальную теорию значения» [14, с. 27] и для машин, и для людей. Другими словами, Уивер пытается сопоставить коммуникацию людей с механической моделью, в которой семантические проблемы, как предполагается, решены.

При введении технической модели в общую теорию коммуникации Уивер предлагает, чтобы проблема эффективности касалась *эффектов* коммуникации: «Ясно, что коммуникация либо влияет на поведение, либо она совсем не имеет никакого отличительного и вероятного эффекта» [14, с. 5]. Тем не менее, Уивер не объясняет, как тождественность или соответствие значения достигается между отправителем и получателем. Понимание значений в модели должно предполагаться, как если коммуникация людей эквивалентна передаче сообщений между машинами. Другими словами, семантические проблемы в коммуникации людей выходят за рамки вопроса.

По мнению Уивера, коммуникация людей становится предметом «всех *процедур*, посредством которых один ум может влиять на другой» [14, с. 3]. Следовательно, понимание как продукт успешной передачи и эффективности коммуникации определяется тем, делает ли слушатель точно то, что ему сказал отправитель. В модели коммуникации Уивера получатель не ведет переговоров и не может их вести, но следует артикуляции «желаемого поведения» отправителя [14, с. 5].

По мнению Уивера, коммуникация не является диалоговой, а процедурной и однонаправленной, ведущей от самоочевидного намерения к самоочевидному поведению. Успешная коммуникация понимается результатом и следствием, а не обсуждением значений.

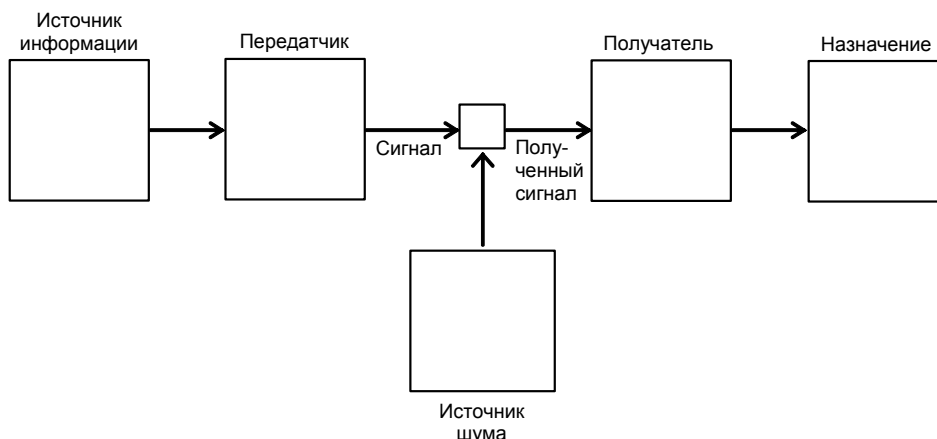


Рис. Графическое представление информационной теории

## Информация в информационной теории

Уивер соглашается, что вкладом теории Шеннона является «действительно основная теория криптографии», однако он утверждает, что эта теория также «вносит вклад в проблему перевода с одного языка на другой, хотя полная картина здесь явно требует рассмотрения значения, так же как и информации» [14, с. 25]. Введение Уивером технических проблем в общую теорию коммуникации включает объяснение некоторых понятий, и одним таким важным понятием является «информация». Как мы увидим далее, Уивер обсуждает «информацию» в различных значениях, некоторые соответствуют понятию информации Шеннона, другие – нет.

Понятие информации в информационной теории Шеннона имеет мало общего со значением быть проинформированным или иметь значение, как принято в нашем общем использовании этого термина. Уивер [14, с. 8] утверждает, что информация «используется в особом смысле, который не следует путать с его первоначальным использованием. В частности, *информацию* не надо путать со значением». Чтобы прояснить эту точку зрения, он добавляет, что «два сообщения, одно из которых сильно нагружено значением, а другое является чистым nonsensom, могут быть совершенно эквивалентными исходя из представленной точки зрения, что касается информации» [14, с.8].

Это вышеупомянутое значение информации соответствует его техническому использованию в эссе Шеннона, в котором информация является алгоритмом максимизации способности передавать сообщение, учитывая возможность канала. Уивер объясняет: «количество информации определяется в самых простых случаях измерением логарифма числа доступных выборов» [14, с.9]. Хотя это значение информации полезно для технической передачи сообщений, оно имеет сомнительное сходство с коммуникацией людей в описании Уивера. В частности, денотаты и коннотации терминов, таких как «информация» и «неопределенность», радикально отличаются в естественных науках и инженерных дискурсах от их общих употреблений.

На основе математической модели разработано кодирование, чтобы контролировать информацию для снижения неопределенности во время передачи. Необходимое кодирование зависит от способности канала связи. По существу, статистические характеристики сигнала должны быть оптимальными для мощности канала связи. Применение того же понимания к коммуникации людей и языку приводит, тем не менее, к механизированной модели связи, которая отрицает саму природу человеческого языка, т.е. то, что почти каждое слово несет множество значений, не говоря уже о небуквальном использовании слов в аналогиях, метафорах или шутках [22].

Аналогия машинной и человеческой коммуникации в тексте Уивера приводит к сомнительной интерпретации понятия информация. Когда он утверждает, что «неопределенность» в информационных системах, порождаемая информацией, имеет хорошие коннотации по сравнению с неопределенностью, созданной шумом, которая имеет плохие коннотации, термины «неопределенность», «хороший», «плохой» и «информация» обычно неправильно интерпретируются как психологические и моральные термины, несмотря на то, что технически это означает производить точное и сжатое кодирование в технической системе.

Использование технической модели для коммуникации людей и согласования словарей, применяе-

мых в тексте Уивера, приводит к непониманию понятия информации в информационной теории с точки зрения намеренной или объективной причины эффектов – первое вызывает желаемое поведение людей, а другое создает психологическую неопределенность, которая радикально отличается от строго определенной вероятной меры в ее оригинальном математическом выражении. Это незаконное присвоение информационной теории Шеннона и получающаяся в результате концептуальная неоднозначность слова «информация» приводят к неправильному пониманию того, что «информация», а не люди, является активным агентом, который приводит к многообразию информационного поведения и событий.

Более того, искаженная модель коммуникации людей в тексте Уивера также приводит к концептуализации человеческой коммуникации и пониманию в соответствии с «поведенческой метафорой канала» (*conduit metaphor*) ([23]; термин взят из [24]). Как мы помним, семантические проблемы широко отрицаются в описании Уивером информационной теории. Тем не менее, большинство коммуникационных или информационных видов деятельности, которые встречаются в библиотеках и информационных системах, включает взаимодействия человек-человек или человек-система. Разработка информационных систем и обучение информационных специалистов зависят от логичной (здравой) теории связи, которая описывает и объясняет, как люди реально взаимодействуют друг с другом и как значения конструируются через акты понимания.

Вкратце, предложение общей теории коммуникации Уивером даёт искажённую картину коммуникации людей, следуя присущим гипотезам. Во-первых, коммуникация является односторонним процессом; диалог отсутствует в модели Уивера. Во-вторых, тождественность значения предполагается или принимается как должное. И наконец, коммуникация понимается как инструментальный или стратегический акт, в том смысле, что его понимание или ситуационное значение является продуктом предполагаемого осуществления причинных намерений, поддерживаемых поведенческими эффектами. Что касается неправильно построенного понятия информации, общая теория коммуникации Уивера отрицает сложности человеческой коммуникации.

## МИР 3 ПОППЕРА

В 1980-1981 гг. Б.С. Брукс опубликовал серию из четырех статей под названием «Основы информатики». В этих статьях Брукс описывает информатику как академическую дисциплину, во-первых, проводя линию между информатикой и библиотековедением, и во-вторых, показывая, как информатика может и должна стать «научной». Статьи Брукса, и в особенности первая часть – («философские аспекты»), оказали влияние на информатику и рассматривались как один из фундаментальных трудов в информатике\*, хотя была также критика статей Брукса, в частности, его интерпретации теории трех миров Поппера, и нужна ли действительно теория Поппера дискурсу библиотековедения и информатики ([см., например, 27]). Сосредоточимся на понятии информации, предложенном и примененном в обсуждении Брукса «философских аспектов».

\* См., например, Белкин [25] и определение информатики Сарацевича [26].

Часть 1 из серии «Основы информатики» является обсуждением философских и теоретических вопросов, в котором Брукс стремится обосновать информатику как научную дисциплину. Брукс [15] утверждает, что туманность информатики как дисциплины и как науки была обусловлена отсутствием теоретических и практических основ. Его главным аргументом является то, что информатика должна рассматриваться как «наука» в том смысле, что феномены должны анализироваться «объективным» способом, который обычно означает, что феномены должны наблюдаться множеством лиц и могут измеряться количественно. Цель признания информатики как объективной науки преобразуется в следующую миссию: «Из всех социальных наук информатика наиболее тесно связана с взаимодействиями между умственным и физическим процессом или между субъективным и объективным способом мышления» [15, с. 126]. Брукс предполагает, что теория Поппера о трех мирах может рассматриваться как начальная точка построения теоретической основы для информатики, в частности в построении фундаментального понимания информации и знания. Он объясняет три мира Поппера:

- Мир 1: Физический мир, космос, в котором Земля, по своей сути, важная для нас, является только незначительной точкой в необъятной вселенной изучения и материи.
- Мир 2: Мир субъективного человеческого знания или «ментальных состояний».
- Мир 3: Мир объективного знания, продуктов человеческого ума, отраженных в языках, искусствах, науках, технологиях – во всех артефактах, которые люди сохранили или распространили по всей Земле.

В понимании Брукса «объективное знание» является знанием, заложенным в артефактах, таких как документы, музыка, искусства и технологии. Знание объективно, так как артефакты могут быть доступны, а их «контенты» оцениваются множеством людей. Брукс выступает против теорий знания, которые являются «субъективистскими». Его объяснение состоит в том, что исследования не могут проводиться объективно, когда знание хранится субъективно в умах людей. По мнению Брукса, знание (и информация) может изучаться только тогда, когда оно «предметно», т.е. вложено в физические, доступные формы. В интерпретации Брукса, например, содержание в книге является продуктом человеческих мыслей и, что важнее, является «объективным знанием» в том смысле, что оно «доступно для всех, кто хочет его изучать» [15, с. 128]. Далее он заявляет, что «фундаментальные сущности Мира 2 и 3, такие же фундаментальные для этих миров, как материя и энергия для Мира 2, являются информацией и знанием» [15, с. 132]. Важно отметить эту аналогию информации и знания для материи и энергии, так как она является аналогией феноменов в «субъективном» мире (Мир 2) и культурном и социальном мирах (Мир 3), что касается естественного и физического мира, который помогает нам понять эмпирическое понятие информации Брукса, применяемое в модели взаимодействий между Миром 2 и Миром 3.

Брукс [15] утверждает, что, хотя практическая работа библиотечных и информационных специалистов является «сбором и организацией записей Мира 3» [15, с. 128], это не является теоретической задачей информатики, так как сама организация документов не дает информатике статус науки и так как «не существует общих предположений, неявных или явных, которые могли бы рассматриваться как ее теоретические основы» [с. 125]. По мнению Брукса, теоретическая задача – «описать и

объяснить их (Мир 2 и Мир 3), если их можно объяснить, и таким образом помочь организовать скорее знание, чем документы для более эффективного использования» [15, с. 128].

### Информация в Мире 3 Поппера

По мнению Брукса, «научная» дисциплина признается используемыми исследовательскими методами, в частности, мнением, что истина предположений может объясняться только путем наблюдений и экспериментальных и количественных методов. Его аналогия материи и энергии для информации и знания предполагает, что информация должна быть наблюдаемым объектом или гипотетически наблюдаемым объектом. «Информация» относится к «продуктам человеческой мысли» и концептуализируется как объективные вещи – содержимое – вложенные в физические формы.

Взаимодействия между Миром 2 и Миром 3 представлены в псевдоматематическом выражении, функции информации:  $K[S] + \Delta I = K[S + \Delta S]$  [15, с. 131]. Это равенство показывает как структуры знания в умах людей могут быть модифицированы в результате изменений в «информации». Равенство означает, что «структура знания» в умах людей может модифицироваться за счет роста или снижения количества мыслей человека, вложенных в артефакты. Это происходит так, как если бы ум человека – так называемая «структура знания» – не играл роли в познании.

Важнее то, что равенство подразумевает, что информация может активно взаимодействовать со структурами знания (Мир 2) с тем, чтобы вызывать изменение в «структуре знания». «Информация» концептуализируется как тип причинного объекта для изменения человеческого ума. Приведение познания к равенству и причинная связь между информацией и ментальными состояниями скрывает от нас то, как люди действительно *обучаются* через чтение книг, слушание музыки и другие «ментальные» или умственные виды деятельности. Ведь обучение – это такая же деятельность, как и виды деятельности быть информированным и знающим, и поэтому они не могут быть изучены как причинная связь между двумя номинальными или псевдоэмпирическими объектами. Скорее, состояния деятельности быть информированным и знающим должны быть изучены как ситуационные виды деятельности, как когнитивные, культурные и социальные феномены одновременно.

Значение информации в интерпретации Бруксом Мира 3 Поппера отличается от модели Шеннона-Уивера. По мнению Брукса, информация относится к «продуктам человеческой мысли», вложенным в физические формы, такие как книги и другого рода артефакты, не как мера, которая имеет место в модели Шеннона-Уивера. «Взаимодействия» между Миром 2 и Миром 3, тем не менее, также концептуализируются как односторонний причинный процесс, т.е. информация изменяет умы людей, а не наоборот. Это должно напомнить нам о процедурной модели коммуникации, а также акценте на эффективность информации, предложенном Уивером.

### МОДЕЛЬ «ДАННЫЕ – ИНФОРМАЦИЯ – ЗНАНИЯ – МУДРОСТЬ» (ДИЗМ)

Третье и последнее значение информации, обсуждаемое в этой статье, является моделью ДИЗМ. Хотя исследователи в области управления знанием обычно рассматривают президентское обращение Р. А. Акоффа к Международному обществу по общим системным исследованиям (International Society for general system re-

search) в 1998 г. как источник модели ДИЗМ [28, 29], считается, что эта модель, по крайней мере, относится к 1960-м гг., хотя, вероятно, без высокого уровня «мудрости». В 2007 г. Зинс [30] опубликовал исследование определений данных, информации, знания и мудрости, выдвинутых исследователями по всему миру в области информатики. Оно показывает, что несмотря на противоречивые определения этих терминов, оно является общепринятым как одно из наиболее хорошо известных определений моделей в информатике.

Модель ДИЗМ часто иллюстрируется как треугольная форма с данными внизу и мудростью наверху. Шедрофф [17] представляет модель как перекрывающиеся круги, но она (модель) также представляет похожий процесс обучения или понимания. Несмотря на формы и размеры модели ДИЗМ, она обычно рассматривается как представление прогрессии от данных к мудрости в умах людей. Акофф [16, с. 3] пишет, «мудрость располагается на вершине иерархии типов, типов содержания человеческого ума. Нисходящими от мудрости являются понимание, знание, информация, а внизу данные». Это представление отражает человеческую мысль и обучение как события, обрабатывающие данные. Репрезентативная точка зрения о знании рассматривает мышление и обучение как количественные события и, таким образом, предполагает, что они могут быть эмпирически исследованы по существу: «Данные являются символами, которые представляют свойства объектов, событий и их окружения. Они являются продуктами *наблюдения*. Наблюдать - значит осмысливать» [16, с. 3]. Он затем добавляет, что «информация, как отмечается, извлекается из данных путем анализа, по многим аспектам которого компьютеры являются экспертами» [16, с. 3].

Модель ДИЗМ, полезная для размышления о потоке данных в обработке информации машинами, рисует ограниченную точку зрения на обучение людей. Как отмечает Махлуп [31, с. 647], когда мы говорим о данных в информатике, мы в основном касаемся «вещей, введенных в компьютер». Он уместно указывает на саму слабость модели ДИЗМ: «Информацию чувствуют, по крайней мере, два человека; один, который говорит (путем выражения, письма, печати, наводки, сигнала), и другой, который слушает, читает, смотрит» [31, с. 645].

В противоположность модели обработки данных, многие будут утверждать, что для людей информация и знания являются необработанными данными; скорее, мы учимся, находясь внутри ситуации и понимая сложные сети связей людей, событий, социальных и политических структур и многие другие вещи. Более того, мы ищем согласия друг с другом относительно того, что вещи означают, через обучающие социальные и культурные средства и категории. Обучение человека не является идентичным обработке данных, так как люди эволюционно учатся через коммуникацию, используя социальные и культурные достижения, тогда как машины в большей степени обрабатывают данные в соответствии с программными процессами. Аналогия данных человеческим стимулам, а результата машинной обработки знанию является странной аналогией, которая озадачивает больше, чем очевидное различие между разработанными и природными агентами. Снова, как, например, с моделью Брукса и с присвоением библиотечковедением и информатикой информационной теории Шеннона через текст Уивера, мы должны задать серьезные вопросы о сравнительно недавних концептуальных историях, которые сформировали такие убеждения, теории и модели.

## Информация в модели ДИЗМ

Значение информации в модели ДИЗМ нуждается в небольшом пояснении. Информация просто понимается как значимые или иначе «обработанные» данные. Акофф [16] заявляет, что, так как информационные системы «генерируют, хранят, ищут и обрабатывают данные,... информация выводится из данных» [16, с. 3]. Значение информации Акофф и многие другие, как цитируется в статье Зинса [30], очевидно основывается на том, как информационные системы работают, но затем этот системный взгляд на информацию используется как эпистемологическая модель для понимания работы человеческого разума и обучения. Эта аналогия машинной обработки и человеческой мысли и коммуникации является еще одним примером одного из двух фундаментальных потоков, которые мешают выражению информационной теории Уивера и пониманию Бруксом теории трех миров Поппера. Другая фундаментальная ошибка, в основном причинного суждения, является неявной не как вопрос метода, а как вопрос эпистемологии: данные являются «причиной» для информации, информация является «причиной» для знания, а знание является «причиной» для мудрости. Причинная связь, напоминающая нам о неявных гипотезах «причинности», придает модели подобие науки как в описании общей теории связи Уивера [14], так и в описании эмпирической когнитивной модели Брукса [15].

Кроме этого, самое очевидное сходство между моделью Шеннона-Уивера и моделью ДИЗМ является их средоточие на эффективности. Акофф делает его «решающим моментом»: «Мудрость есть способность *увеличить эффективность*» [16, с. 5]. По мнению Акоффа, модель ДИЗМ демонстрирует рост в эффективности на каждом из ее уровней. Треугольная форма, в которой модель ДИЗМ часто показывается, предполагает, что имеется консолидация мощного или потенциального «эффекта» от низших к высшим уровням. Инверсная логика этого является такой, что чем большими данными обладает один человек, тем больше информации, знания и мудрости он может порождать для улучшения эффективности. Знание и мудрость, таким образом, становятся понятными как продукты обработки данных, а не обучения и опыта. Более того, понимание объясняется в терминах телеологических процессов систем обработки и развития: «Увеличить эффективность – это либо увеличить вероятность производства желаемого результата при фиксированных ресурсах, либо снизить количество ресурсов, требуемых для его производства с особой вероятностью» [16, с. 4]. Здесь снова модель ДИЗМ имеет сходство с техническим снижением для понимания и коммуникации, т.е. критерием описания Уивером информационной теории Шеннона.

Однако имеются онтологические различия между двумя моделями: модель Шеннона-Уивера рисует модель коммуникации, тогда как модель ДИЗМ рисует теорию знания, отображающую традицию эмпиризма и метафорически полагающуюся скорее на обработку данных, чем на коммуникативную модель для отображения ее понятия информации. Модель ДИЗМ полагается в большей степени на восприятие смысла и сбор данных как начальные стадии в производстве знания. Она разделяет предположение эмпирической эпистемологии, что чувственные ощущения являются источником для человеческого понимания. Короче говоря, хотя модель ДИЗМ может хорошо моделировать машинную обработку данных, с точки зрения понимания людей она остается эпистемологически приземленной в версии

эмпиризма с дополнительной технической метафорой. Если мы уберем техническую метафору, окончательное понимание информации должно будет пренебречь исследованием языка, коммуникации и обучения в человеческом бытии в прошедшие три столетия.

## **ОБСУЖДЕНИЕ: ЗНАЧЕНИЯ, ГИПОТЕЗЫ И ПОСЛЕДСТВИЯ ИССЛЕДОВАНИЙ**

Термин «информация» часто используется для представления многообразия вещей, событий или выражений. Другими словами, термин «информация» не может интерпретироваться как вещи, события или выражения сами по себе, если он не находится внутри контекста или не понимается осмысленно. Этот факт становится справедливым в нашем повседневном использовании термина, также как концептуальное построение в трех обсуждаемых теориях. Из-за гибкости в значении возникают многие различные смыслы и коннотации этого термина. Более того, объединение значений не является необычным, и некоторые из них приводят к идеологическому использованию. Например, обычное понимание «информации» в статье Брукса [15] - продукты человеческих мыслей, облеченные в физическую форму - часто несет с собой коннотации измеряемого объекта, как нечто, что может быть закодировано и причиной познания. Неоднозначное объединение различных значений и смыслов, в частности, когда основанное в рамках эмпирических эпистемологий, придает слову «информация» значительную силу не только в профессиональной речи, но и в современной культуре и обществах.

Значения информации в трех теориях, обсуждавшихся выше, отличаются друг от друга. В модели Шеннона-Уивера информация относится к вероятностной мере. Действительно, она, тем не менее, является научным термином, неправильно истолкованным многими, т.е. она относится к важным сообщениям, которые могут изменить поведение людей на основе искаженной модели коммуникации людей. В фундаментальной статье Брукса [15] информация понимается как человеческие мысли, вложенные в физические формы. В модели ДИЗМ информация понимается как организованные и обработанные данные. Несмотря на различные значения, концептуализации информации являются очень схожими в терминах эмпирических эпистемологических обязательств, которые соответствуют им.

В трех обсуждавшихся теориях информация понимается как предметы или единицы, которые могут объективно измеряться (модель Шеннона-Уивера), наблюдаться («информация» Брукса) или обрабатываться (модель ДИЗМ). Следовательно, предполагается, что феномены информации должны измеряться в терминах причинных эффектов. Другими словами, некоторое понятие физической или логической причинности охватывает человеческую прагматику понимания и коммуникации так, чтобы последние могли быть поняты «научно», а значит, дисциплины, которые их понимают, должны быть научными с соответствующими методами [32].

Социальным последствием последней точки зрения является то, что социальная «эффективность» становится понятной детерминистически и, более того, через неявные или явные машинные метафоры, циркулирующие в этих текстах, в терминах телеологии разработанных технологических систем. Обучение и понимание людей понятны как если бы они были аналогичными технологическим системам и обработке, а технологические системы могут быть рассмотрены как необходимые или

достаточные помощники этому. Однако в противоположность всему этому широко понимается то, что человеческая коммуникация и познание развиваются в результате основанного на опыте обучения и развития в особой среде с использованием социальных и культурных норм и средств, а также доступных материалов. Понятия «данные», «информация» и т.д., как рассматривалось в обсуждавшихся теориях, а также их детерминистические причинные модели и средства вряд ли исследуют это универсальное понимание.

## **ЗАКЛЮЧЕНИЕ: ЧТО ТАКОЕ ИНФОРМАЦИЯ И ЧТО ДЕЛАЕТ ЕЕ ИНФОРМАЦИЕЙ?**

Одним исследовательским последствием концептуализации информации как «объективной» и как имеющей причинные силы является отрицание человеческого, культурного и социального компонентов информации. Изучает ли исследование процессы информационного поиска, информационное поведение, алгоритмы по улучшению информационного поиска или модели ссылок, один вопрос часто остается без ответа в этих научных исследованиях: что такое информация и что делает ее «информацией»?

Обзор Кейса [33] исследований информационного поведения представляет собой интересный случай. В своем обзоре Кейс категоризирует огромный объем исследований информационного поведения по профессии, социальной роли и демографической группе, но не по «информации». Что ищут люди разных профессий, социальных ролей и демографических групп? Как мы помним, термин «информация» часто используется для представления многообразия вещей. Информация не является самоочевидной.

Бакленд [34] предположил, что «то, что разумно рассматривается или не рассматривается как информация, зависит от согласия, или, по крайней мере, от некоторого консенсуса» [34, с. 357]. Эта ситуационная точка зрения относительно информации разделяется многими в этой области (см., например, [12, 35-36]); однако мы еще задали вопросы о строении и деконтекстуализации информации: какие процессы приводят к консенсусу по поводу того, что является и что не является информацией, и идеологиям, которые могут быть вовлечены в социальный мир [37]? И как использование термина «информация» или присвоение его значений в окружении или текстах, отличных от первоначального значения или концептуализации (описание Уивера информационной теории является таким примером), влияет на информационный дискурс?

С одной стороны, анализы построения информации являются необходимыми для улучшения информационных служб и разработки информационных систем, так как такой анализ прояснит причины того, как что-то считается или не считается информацией, а также причины, которые мешают идентификации, производству и использованию определенных типов информации. Другими словами, что *является* информацией в различных ситуациях и окружениях, процесс ее построения должен быть прозрачным. С другой стороны, анализ деконтекстуализации информации изучает предназначение «информации», которое может быть идеологическим и которое может ускользать от явной осведомленности производителей и пользователей информации. Иначе, такой анализ стремится открыть искаженные и неправильно используемые понятия информации и их причины, применения и последствия.

Другим исследовательским последствием эпистемологических взглядов, соответствующих трем теориям, является бесспорное применение некоторого понятия научного метода. Если информация является чем-то, всегда наблюдаемым и измеряемым, подобно скале в физической среде, тогда исследовательские методы, используемые в физической науке, могут считаться разумными, так как обоснованность такого рода научных исследований может оцениваться любым, кто обладает надлежащим образовательным профилем и подготовкой. Когда Брукс [15] утверждает, что информация и знание должны изучаться, подобно материи и энергии, в физическом мире, он полагает, что информация и знание являются наблюдаемыми, измеряемыми и относительно стабильными во времени и в культурах. Проблема снова состоит в том, что термин «информация» относится к широкому разнообразию вещей и феноменов; интерпретация его значения или объекта ссылки зависит от ситуации и контекста. Поэтому, научные методы, используемые для его исследования, должны выбираться соответственно.

Например, информация может относиться к физическому описанию книги или дискуссии о будущем книг. Природа описания книги подобна природе описания скалы: неизменяемая, обозримая и измеряемая, тогда как дискуссия о будущем книг является преходящей, необозримой и не измеряемой после ее окончания. Более того, факторы, такие как человеческие намерения и язык, культурные нормы, институтские практики, экономические условия и социальные структуры, могут влиять на появление и взаимодействия в дискуссии. Научные методы, которые подходят для изучения сложностей социального мира, требуются для изучения информации в том смысле, что построение информации включает мнения, намерения, желания, а также культурные формы и социальные практики. Методологические анализы, такие как оценка Давенпорта [38] научных методов, обычно используемая в исследовании информационного поведения, в частности, по дуализму наук (т.е. разветвление наук (бифуркация) и гуманитарные науки) и полезность и эффективность научных методов надолго запаздывают в исследовании информатики.

В этой статье обсуждаются значения термина «информация» в модели Шеннона-Уивера, интерпретации модели Поппера Бруксом и модели ДИЗМ и их гипотезы, а так же последствия исследования. Показывается, что понятие информации в этих теориях является таким, что оно представляется количественным объектом или событием, имеющим причинные эффекты на людей. Предполагается, что гипотезы приводят к а) небрежности в культурном и социальном построении информации и б) бесспорному использованию «научных» исследовательских методов. Однако эта статья не предполагает, что научные последствия в ходе следования эмпирической эпистемологии прошли незамеченными со временем. В действительности, различные повороты информатики [39] спокойно отвечают недостаточностям различных эпистемологических парадигм. Сфокусированное на системе или пользователе исследование в научной информации или повседневном поиске информации всеми стремится улучшить проектирование информационной системы и продвигать позитивное социальное изменение. Что прошло незамеченным — это эпистемологические гипотезы фундаментальных понятий и их связи с исследовательскими методами. В этой статье изучались понятия информации в трех фундаментальных теориях и обсуждались их гипотезы и

последствия исследований в надежде открыть эпистемологическую и методологическую дискуссии в будущем.

*Благодарность.* Автор благодарит Рона Дея и рецензентов за их конструктивные предложения и комментарии.

## ЛИТЕРАТУРА

1. Capurro R., Hjørland B. The concept of information// Annual Review of Information Science and Technology. — 2003. — Vol. 37. — P. 343–411. — Medford, NJ: Information Today.
2. Day R. E. The modern invention of information: Discourse, history, and power. — Carbondale, IL: Southern Illinois University Press, 2001.
3. Frohmann B. Deflating information: From science studies to documentation. — Toronto, Ontario: University of Toronto Press, 2004.
4. Braman S. Change of state: Information, policy, and power. — Cambridge, MA: The MIT Press, 2009.
5. Floridi L. Information: A very short introduction. — New York: Oxford University Press, 2010.
6. Saab D.J., Riss U.V. Information as ontologization// Journal of the American Society for Information Science and Technology. — 2011. — Vol. 62, No.11. — P. 2236–2246.
7. Marchionini G. Information concepts: From books to cyberspace identities. — Morgan & Claypool, 2010.
8. Gleick J. The information: A history, a theory, a flood. — New York: Pantheon Books, 2011.
9. Bates M.J. Information and knowledge: An evolutionary framework for information science// Information Research. — 2005. — Vol. 10, No. 4.
10. Bates M.J. Fundamental forms of information// Journal of the American Society for Information Science and Technology. — 2006. — Vol. 57, No. 8. — P. 1033–1045.
11. Bates M.J. Hjørland's critique of Bates' work on defining information// Journal of the American Society for Information Science and Technology. — 2008. — Vol. 59, No.5. — P. 842–844.
12. Hjørland B. Information: Objective or subjective/situational?// Journal of the American Society for Information Science and Technology. — 2007. — Vol. 58, No.10. — P. 1448–1456.
13. Hjørland B. The controversy over the concept of “information”: A rejoinder to Professor Bates// Journal of the American Society for Information Science and Technology. — 2009. — Vol. 60, No.3. — P. 643.
14. Shannon C.E., Weaver W. The mathematical theory of communication. — Urbana: The University of Illinois Press, 1964.
15. Brookes B.C. The foundations of information science: Part I philosophical aspects// Journal of Information Science. — 1980. — Vol. 2. — P. 125–133.
16. Ackoff R.L. From data to wisdom: Presidential address to ISGSR, June 1988// Journal of Applied Systems Analysis. — 1989. — Vol. 16. — P. 3–9.
17. Shedroff N. An overview of understanding/ R. S.Wurman (Ed.), Information anxiety 2. — P. 27–29. — Indianapolis, IN: Que, 2001.
18. Greene B. The hidden reality: Parallel universes and the deep laws of the cosmos. — New York: Alfred A. Knopf, 2011.



19. *von Baeyer H.C.* Information: The new language of science. — Cambridge, MA: Harvard University Press, 2003.
20. *Shaw D., Davis C.H.* Entropy and information: A multidisciplinary overview// Journal of the American Society for Information Science. — 1983.— Vol. 34, No. 1.— P. 67–74.
21. *Kalbach J.* On uncertainty in information architecture// Journal of Information Architecture. — 2009. — Vol. 1, No.1. — P. 48–55.
22. *Lakoff G., Johnson M.* Metaphors we live by. — Chicago: The University of Chicago Press, 1980.
23. *Day R.E.* The “conduit metaphor” and the nature and politics of information studies// Journal of the American Society for Information Science. — 2000.—Vol. 51, No. 9. — P. 805–811.
24. *Reddy M.J.* The conduit metaphor: A case of frame conflict in our language about language Metaphor and Thought. — Cambridge: Cambridge University Press, 1993. — p. 164–201.
25. *Belkin N.J.* The cognitive viewpoint in information science//Journal of Information Science. — 1990. — Vol. 16. — P. 11–15.
26. *Saracevic T.* Information science// Journal of the American Society for Information Science. — 1999. — Vol. 50, No.12. — P. 1051–1063.
27. *Rudd D.* Do we really need World III? Information science with or without Popper// Journal of Information Science.—1983. — Vol. 7. — P. 99–105.
28. *Frické M.* The knowledge pyramid: A critique of the DIKW hierarchy// Journal of Information Science. — 2009. — Vol. 35, No.2. — P. 131–142.
29. *Rowley J.* The wisdom hierarchy: representations of the DIKW hierarchy// Journal of Information Science. — 2006. — Vol.33, No.2. — P. 163–180.
30. *Zins C.* Conceptual approaches for defining data, information, and knowledge// Journal of the American Society for Information Science and Technology. — 2007. — Vol. 58, No.4. — P. 479–493.
31. *Machlup F.* Semantic quirks in studies of information/ F. Machlup, U. Mansfield (Eds.), The study of information: Interdisciplinary messages. — P. 641–671. — New York: John Wiley & Sons, 1983.
32. *Day R.E.* Tropes, history, and ethics in professional discourse and information science// Journal of the American Society for Information Science. — 2000. — Vol. 51, No. 5. — P. 469–475.
33. *Case D.O.* Looking for information: A survey of research on information seeking, needs, and behavior (2nd ed.). — Boston: Elsevier, 2007.
34. *Buckland M.K.* Information as thing// Journal of the American Society for Information Science. — 1991. — Vol. 42, No. 5. — P. 351–360.
35. *Eklbia H.R., Evans T.P.* Regimes of information: Land use, management, and policy// The Information Society. — 2009. — Vol. 25. — P. 328–343.
36. *Ma L.* Information as discursive construct// Paper presented at the American Society for Information Science and Technology Annual Meeting, Pittsburgh, PA, October 22–27, 2010.— 2010.
37. *Benoit G.* Critical theory as a foundation for pragmatic information systems design// Information Research. — 2001. — Vol.6, No.2.
38. *Davenport E.* Confessional methods and everyday life information seeking// Annual Review of Information Science and Technology. — 2010. — Vol. 44. — P. 533–562. — Medford, NJ: Information Today, 2010.
39. *Cronin B.* The sociological turn in information science// Journal of Information Science. — 2008. — Vol. 34, No.4. — P. 465–475.

# Пять звезд электронных журнальных статей: подход к оценке статей\*

**Дэвид ШОТТОН**  
(David SHOTTON)

Зоологический факультет Оксфордского университета, Великобритания

*Предлагаются пять факторов — рецензия, открытый доступ, обогащенное содержание, доступные наборы данных и машиночитаемые метаданные — как пять звезд электронных журнальных статей, созвездие пяти независимых критериев в многомерном издательском универсуме, по которым электронные журнальные статьи могут быть оценены, чтобы видеть, насколько они соответствуют современным представлениям об улучшенных научных коммуникациях. Достижение каждой из этих публикационных осей может отличаться, уподобляться другим звездам в созвездии, сияющим с различной светимостью. Предлагается пятизвездная шкала, по которой журнальная статья может оцениваться, и даются представления в виде диаграмм для каждой оценки. Поскольку принятые для каждой из этих шкал критерии являются отчасти произвольными и поскольку оценка определенной статьи на каждой оси может включать элементы субъективного суждения, эти пять звезд электронных журнальных статей обеспечивают концептуальную структуру, позволяющую судить о степени, которой любая статья достигает или до которой падает ниже идеала, полезного для авторов, редакторов и издателей. На примерах таких оценок своих собственных последних публикаций автор этой статьи показывает их соответствие семантической публикации.*

## ВВЕДЕНИЕ

Многие люди, возможно, знакомы с пятью звездами связанных открытых данных, предложенных Бернерсом-Ли, т.е. с прирастающими ступенями, категоризирующими публикацию открытых данных в сети по уровням растущей полезности, которые охватывают существующее распространённое представление о семантической сети как о сетевых связанных открытых данных (Web Linked Open Data) и которые отдельные лица могут использовать для оценки публикации своих собственных данных.

Пять звезд связанных открытых данных Бернерса-Ли таковы [1]:

★ Сделайте ваши данные доступными в сети (в любом формате), но с открытой лицензией, чтобы данные были открытыми.

★★ Сделайте их доступными в виде машиночитаемых структурированных данных (например, используйте программу Excel вместо изображения таблицы).

★★★ Как и предыдущий пункт, но используйте не собственные форматы (например, CSV вместо Excel).

★★★★ Все вышеупомянутые пункты плюс: используйте открытые стандарты из W3C (RDF и SPARQL) для определения вещей, с тем чтобы люди могли указать на ваш материал.

★★★★★ Все вышеупомянутые пункты плюс: присоедините ваши данные к данным других людей, чтобы обеспечить контекст.

В дополнение к этому я хочу предложить пять звезд электронных журнальных статей, в частности, чтобы охарактеризовать потенциал для усовершенствования первичных средств научной коммуникации, ставших возможными благодаря сетевым технологиям. Основой для этих пяти звезд электронных журнальных статей служат семантическое издательство (семантическая публикация), принципы будущих научных коммуникаций и сама семантическая сеть.

Тогда как сторонники семантической сети иногда пытались походить на пророков Ветхого завета, чьи сообщения истины прошли не замеченными всем народом, сообразительность ряда влиятельных сторон, таких как BBC, и умелая спекуляция понятиями семантической сети под знаменем «связанных данных», недавно принесли технологиям семантической сети более широко-масштабное признание. Принципы совершенно простые. Если предметы и их связи могут быть установлены и определены в машиночитаемой форме путем использования единственного URI (универсального идентификатора ресурсов), ссылающегося на публично доступные и общепринятые структурированные определенные словари (онтологии), и если каждая из этих связей выражена как простое предложение субъект — предикат — объект («тройка»), следуя синтаксису структуры описа-

\* Перевод Shotton D. The five stars of online journal articles — a framework for article evaluation // D-Lib Magazine. - 2012. — Vol. 18, No. 1/2. — <http://www.dlib.org/dlib/january12/shotton/01shotton.html>

ния ресурса (RDF – Resource Description Framework), тогда такие предложения могут быть включены во взаимосвязанные информационные сети (графы RDF), в которых истинное содержание каждого оригинального предложения поддерживается созданием сети знания, т.е. семантической сети.

Онтологические описания предметов и их связей дают возможность данным из независимых источников быть включенными без двусмысленности или потери точности смысла, ситуация, которая не могла бы быть возможной, если бы предметы были описаны другим образом, таким как расширенный язык разметки – XML, где отсутствие общепринятых значений для разметки терминов часто приводит к путанице с синонимами (например, когда «создатель» в одной схеме эквивалентен «композитору» или «хореографу» – в другой), либо с омонимами (например, потенциально различные значения разметки термина ‘gift’ – “дарение”, означающее ‘present’ – “подарок” в английской базе данных, но ‘poison’ – “отраву” в немецкой базе данных).

Сейчас существует много достойных примеров того, как использовать технологии семантической сети «под покровом», что позволяет объединять в унифицированные и связанные услуги первоначально закодированные данные, используя модели несовместимых метаданных, и данные, находящиеся в неоднородных базах данных. Лучший пример, в который я лично был вовлечен, – это база данных CLAROS, «Мир искусства в семантической сети», в которой информация, описывающая объекты античного искусства, находящиеся в мировых музеях, включена из ряда научных источников [2]. Достоинства технологий семантической сети для библиотек недавно обсуждались на ежегодном совещании «Семантические сети в библиотеках» под названием «Научная коммуникация в сети данных», состоявшемся в Гамбурге (Германия) в ноябре 2011 г.

## Семантическое издательство

Журнальная публикация как первичный канал распространения и публичного представления новых научных результатов является жизненно важным компонентом научного рабочего потока, а ее ключевой товар, оригинальная научная статья, представляет первостепенную важность, поскольку она обеспечивает датированную «версию записи» авторских гипотез, подтверждающую результаты и выводы на время публикации, подкрепленные рецензией, и как таковая становится неизменной частью научного отчета. Основной формат научной журнальной статьи мало изменился с момента его принятия около 350 лет назад. Он остается линейным риторическим повествованием, в котором авторы пытаются убедить читателей в правильности конкретных гипотез, представляя экспериментальные свидетельства, взятые из больших массивов данных.

В настоящее время большинство журнальных издателей использует Интернет просто как удобный механизм для распространения журнальных статей в формате PDF, обеспечивающем электронные копии печатных страниц. Поскольку документы в формате PDF являются подходящими для печати и автономного чтения, типичный недостаток любой формы семантического обогащения или взаимодействия пользователя, а также трудности, которые они представляют для машинной интерпретации, в настоящее время препятствуют разработке автоматизированных услуг, которые должны обо-

гатить содержание журнальных статей или связать информацию между статьями.

Тем не менее, в последнее время предпринимаются различные инициативы, чтобы изменить это положение, изучающие, как сеть может использоваться для обогащения электронных научных коммуникаций различными способами, которые невозможны в печатной форме. Например, были сделаны экземпляры, порождающие семантические обогащения версий HTML журнальных статей [3,4], были созданы сетевые услуги по извлечению из текста, которые могут автоматически добавлять семантические метки названным статьям в тексте HTML [5] или извлекать контекстуальную информацию из цитируемых статей [6], и «толковые» программы для чтения документов PDF, такие как Utopia Documents, должны быть разработаны, чтобы обеспечить обложки с аннотациями для обогащения другого статичного содержания статей в формате PDF [7]. Сильвио Перони и автор настоящей статьи разработали онтологию семантического издательства и ссылок (SPAR – Semantic Publishing and Referencing) для обеспечения таких разработок [8,9], и издатели, включая Project Prospect Королевского химического общества, Article of the Future и Pensoft Journals корпорации Elsevier, начали обеспечивать семантически обогащенные журнальные статьи как часть своих обычных издательских рабочих потоков.

Публикация журнальных статей с такими добавлениями становится известной как «семантическое издательство», термин, который я определил как использование технологий простой и семантической сетей для того, чтобы:

- обогатить содержание электронной научной статьи, например средствами интерактивных рисунков, перепорядоченных библиографических списков, семантических линз (показ графов, когда щелкаете мышью на таблицу числовых данных, или анимации, когда щелкаете мышью на диаграммы);

- повысить значение содержания статьи, например средствами семантической разметки названных сущностей со связями к описательным определениям значений терминов и понятий и к дополнительной информации, касающейся этих сущностей (скажем, к существующей базе данных по протеинам от определенных названий протеинов);

- обеспечить связи к другим информационным источникам, релевантным этой статье, например к домашним страницам авторов, сетевым сайтам поставщиков реагентов и соответствующим международным организациям (скажем, к ВОЗ от эпидемиологической статьи и наоборот);

- обеспечить *прямые* связи со всеми цитируемыми ссылками статьи;

- обеспечить доступ к данным в опубликованной статье в действенной форме, например к таким, как импортируемая электронная таблица или файл CSV;

- связать статьи с полными научными наборами данных, которые поддерживают статьи;

- облегчить интеграцию данных статьи с семантически связанной научной информацией в другом месте – в литературе или сети;

- помочь раскрытию (содержания статьи) путем открыто публикуемых машиночитаемых описательных метаданных, например путем детализации библиографического описания статьи, резюмирования ее

содержания и обеспечения библиографических элементов цитируемых ею ссылок.

Цель такого семантического издательства состоит в том, чтобы данные, информация и знания, описанные в электронной статье, могли стать легче обнаруживаемыми, извлекаемыми, включаемыми и используемыми для многих целей.

### Будущее научной коммуникации

В 2011 г. состоялись четыре важных совещания, собравшие вместе ученых, компьютерных специалистов и научных издателей, чтобы обсудить будущее научной коммуникации. Первый из них, семинар, названный «Что скрывается за PDF», организованный и проведенный в январе 2011 г. Филом Баурне в Калифорнийском университете (г. Сан-Диего), строился на основе более раннего семинара HyPER, организованного в мае 2010 г. в Амстердаме Анитой де Ваард из Elsevier Labs [10]. Следующим было совещание «Что скрывается за фактором влияния», организованное Камероном Нейлоном из STFC (Science and Technology Facilities Council) в штаб-квартире Wellcome Trust в Лондоне в мае 2011 г., на котором рассматривались альтернативные оценки (параметры, критерии) фактора журнального влияния для оценки научных заслуг, в частности, ученого. В августе 2011 г. состоялось следующее совещание «Будущее научной коммуникации», организованное Филом Баурне из Калифорнийского университета в г. Сан-Диего, Тимом Кларком из Гарвардского университета, Робертом Дейлом из Университета Маквасери (Австралия), Анитой де Ваард из Elsevier Labs., Айвеном Германом из W3C, Эдуардом Хоуви из Университета шт. Южная Каролина и автором настоящей статьи, которое прошло в Германии как семинар Научного центра по информатике «Schloss Dagstuhl Perspectives Workshop». Это привело к формированию сообщества Force 11 Community, предназначенного для совершенствования научной коммуникации и электронной эрудиции, и публикации в октябре 2011 г. отчета «Force 11 White Paper» [11], который являлся свидетельством как для проекта «Наука как общественное предприятие» Королевского общества, так и для общественной консультации «Делание реальностью открытых данных» кабинета министров Великобритании. Совсем недавно, в октябре 2011 г., корпорация Microsoft Research и Гарвардский университет провели совместное совещание в г. Кембридж (шт. Массачусетс, США), названное «Преобразование научной коммуникации», которое продвинуло эти идеи вперед. Мысли, озвученные на этих совещаниях, внесли значительный вклад в формирование пяти звезд электронных журнальных статей.

### ПЯТЬ ЗВЕЗД ЭЛЕКТРОННЫХ ЖУРНАЛЬНЫХ СТАТЕЙ

Я предлагаю пять факторов – рецензия, открытый доступ, обогащенное содержание, доступные наборы данных и машиночитаемые метаданные – как пять звезд электронных журнальных статей, созвездие пяти независимых критериев в многомерном издательском универсуме, по которым электронные журнальные статьи могут

быть оценены и которые предназначены для характеристики потенциала совершенствования журнальной статьи, возможного с помощью сетевых технологий.

Пять звезд электронных журнальных статей



#### Рецензия

Обеспечьте рецензирование вашей статьи, чтобы гарантировать ее научную значимость, качество и целостность.

#### Открытый доступ

Обеспечьте другим бесплатный открытый доступ для чтения и многоцелевого использования вашей публикуемой статьи, чтобы гарантировать ей наиболее возможную читаемость и используемость.

#### Обогащенное содержание

Используйте полный потенциал сетевых технологий и сетевых стандартов, чтобы обеспечить интерактивность и семантическое обогащение содержания вашей электронной статьи.

#### Доступные наборы данных

Сделайте так, чтобы все данные, обеспечивающие результаты вашей статьи, публиковались по открытой лицензии с достаточными метаданными для возможности их переинтерпретации и многоцелевого использования.

#### Машиночитаемые метаданные

Публикуйте машиночитаемые метаданные, описывающие вашу статью, и цитируемые вами ссылки так, чтобы эти описания могли быть открываемыми и используемыми для многих целей в автоматическом режиме.

Тогда как пять звезд связанных открытых данных Тима Бернерса-Ли строились одна за другой, представляя степени достижения или полноты вдоль одной единственной оси электронной публикации данных, предлагаемые пять звезд электронных журнальных статей являются дополняющими, формируя созвездие, расположенное на пяти независимых осях в многомерном издательском универсуме, каждая из которых может оцениваться по собственным критериям. Конечно, степень достижения на каждой из этих издательских осей может варьироваться, равняться различным звездам в созвездии, сияющим с различной светимостью.

Таким образом, пять звезд электронных журнальных статей дают более богатое представление. Каждая звезда весьма желательна в своем собственном роде, однако богатое представление достигается только комбинацией всех их, к чему мы должны действительно продвигать научную коммуникацию. Давайте теперь рассмотрим, как мы можем оценивать эффективность отдельных статей по каждой звезде. Мои комментарии предназначены в первую очередь для авторов, но каждому должно быть ясно, что реализация этих издательских целей потребует активного и полного энтузиазма сотрудничества издателей и редакторов журналов.

## Рецензия

**Обеспечьте вашей статье рецензирование, чтобы гарантировать ее научную ценность, качество и целостность.**

Гарантия качества журнальных статей традиционно обеспечивалась анонимным предпубликационным рецензированием. В моей личной практике рецензирование всегда имело позитивный опыт, рецензии были справедливыми, и исправления вносились в ответ на замечания рецензентов, постоянно улучшая общую читаемость и качество статей. Однако такая практика с течением времени подверглась серьезному сомнению по нескольким причинам. Во-первых, система, практикующая предпубликационное рецензирование, является неэффективной и медлительной, работающей под давлением все растущего числа статей, подлежащих публикации. Во-вторых, ученые, которые желают взять на себя эту деятельность для пользы своих коллег-ученых и безвозмездно со стороны научных издателей, в значительной степени неохотно делают это, поскольку хорошее рецензирование требует усилия, за которое рецензенты получают скудное вознаграждение в виде научного признания, когда они находятся под давлением со стороны других лиц. В-третьих, эта услуга критикуется за неудачу достичь своей цели гарантирования того, что эти статьи, принятые к публикации, всегда являются высокого качества. Наконец, иногда в рамках представленной компетенции встречается крайне плохая научная практика, позволяющая рецензенту задерживать публикацию работы конкурента, поскольку проводимое исследование основано на украденных идеях, таким образом, обеспечивая рецензенту научное доверие, которое справедливо принадлежит авторам рецензируемой статьи.

Три подхода предлагаются для совершенствования процесса рецензирования и гарантии против такой плохой практики. Во-первых, анонимность рецензента должна быть отменена, чтобы не только уменьшить нарушения, но и чтобы позволить научному признанию быть вознагражденным более прозрачно для большинства хороших рецензентов, которые посвящают свое время этому процессу. Во-вторых, рецензии должны публиковаться вместе с рецензируемой статьей, чтобы читатели могли видеть вклад, внесенный рецензентами в окончательный текст. В-третьих, наиболее противоречивый подход, процесс гарантии качества должен быть отделен от акта публикации.

Небольшое число журналов сейчас приняло полностью открытое рецензирование для всех своих статей с очевидно удовлетворительными результатами. Однако критики этой политики указывают, что, по крайней мере, в некоторых дисциплинах, таких как гуманитарные науки, которые имеют менее открыто критикуемые культуры, отсутствие анонимности может удерживать рецензентов от выдвижения справедливых замечаний, таким образом, ослабляя процесс рецензирования.

Поскольку публикация сейчас может быть предпринята полностью в электронном режиме за значительно меньшую стоимость по сравнению с печатью статей в традиционных журналах, то не требуется, чтобы статья была подготовлена в ее окончательной форме до ее публикации. Постпубликационное рецензирование позволяет расширить ответственность за рецензирование от двух-трех отдельных ученых, выбранных редакционным персоналом журнала, до широкого научного сообщества, чья обратная связь может затем быть включена в измененную версию статьи, которая затем публикуется.

Такое постпубликационное рецензирование считается многими учеными «легковесным» и критикуется как хорошо работающее в пользу спорных или очень интересных статей, но менее эффективно для обоснованных статей более ограниченного интереса, не в меньшей степени из-за того, что читатели находятся под своим собственным давлением нехватки времени и неохотно занимаются деятельностью, для которой не установлены механизмы научного вознаграждения. Однако постпубликационное рецензирование является строгой нормой для тех, кто публикует спецификации Интернета и сетевые документы по стандартам – RFC (Requests for Comments – требования для комментариев), издаваемые IETF (Internet Engineering Task Force), и предлагаемые рекомендации, издаваемые консорциумом W3C (World Wide Web Consortium), – где всё намерение начальной публикации состоит в том, чтобы сделать эти документы доступными для постпубликационного рецензирования за определенный период времени, в течение которого замечания и критика от любых заинтересованных сторон получались и использовались, прежде чем новые стандарты формально согласованы и опубликованы.

Статьи, публикуемые в электронном режиме, могут дополнительно получать замечания и быть предметом одобрительных оценок от читателей, что касается качества статьи, будучи определенным, по меньшей мере, в части полученной от нее пользы, хотя на практике иметь возможность делать такие замечания по публикуемым статьям ограничено.

Эти альтернативные возможности позволяют нам оценивать процесс рецензирования, первый из пяти звезд электронных журнальных статей, в терминах его эффективности и открытости, используя следующую простую пятибалльную шкалу – от 0 до 4:

### 0 – Рецензия отсутствует

Статья публикуется без предпубликационного рецензирования, например в *Nature Proceedings* или на сервере препринтов, таком как *arXiv*.

### 1 – Предпубликационная рецензия

Статья прошла критическую оценку и рецензирование двумя или более соответствующими экспертами по инициативе редактора журнала и была принята к публикации на основе замечаний и предложений рецензентов. Рецензенты остаются анонимными и их рецензии не публикуются. Эта ситуация характерна для многих журнальных статей. Некоторые журналы, например *PLoS ONE*, требуют от рецензентов оценивать не потенциальную важность, влияние или широту привлекательности статьи, а только ее научную обоснованность.

### 2 – Ответная рецензия

Статья критикуется и рецензируется двумя или более соответствующими экспертами (либо до, либо после начальной публикации), и автор должен существенно отреагировать на эти замечания путем возражения на критику рецензентов, посланную редактору журнала, или внесением изменения в окончательно публикуемую версию статьи в свете замечаний и предложений рецензентов. Рецензенты остаются анонимными, и их рецензии не публикуются. Большинство журналов принимают эту политику для своих статей.

### 3 – Постпубликационная рецензия

В дополнение к ответному рецензированию журнальной статьи читатели могут давать посткомментарии после публикации. Они могут продолжить дискуссию в статье, привлечь внимание к дополнительным соответствующим результатам или потребовать авторских вы-

водов. Эти замечания доступны другим читателям, а автор оповещается и может ответить на них, таким образом складывается постпубликационная рецензия. Журналы *PLoS (Public Library of Science)*, среди прочих, обеспечивают это для своих статей. Другие формы публикационного рецензирования также возможны, как, например, в RCF и Candidate Recommendations.

#### 4 – Открытая рецензия

Весь процесс рецензирования полностью прозрачен. Каждая поступившая рукопись немедленно делается доступной на сетевом сайте журнала. Рецензии и замечания от читателей приветствуются и рассматриваются наряду с аналогичными формальными рецензиями, запрашиваемыми от экспертов журналом. Все рецензии, авторские ответы и начальная и окончательная версии статьи публикуются, а соответствующим рецензентам и редакторам выражается благодарность именно в окончательной версии статьи. Журнал *BMJ Open* практикует такое открытое рецензирование, а журнал *Semantic Web Journal* очень поддерживает его, в то же время предоставляя рецензентам выбор оставаться анонимными.

### Открытый доступ

**Обеспечьте другим бесплатный открытый доступ, как для чтения, так и для многоцелевого использования вашей публикуемой статьи, чтобы гарантировать ей наибольшую возможную читательскую аудиторию и использование.**

Наиболее фундаментальное изменение, которое Интернет внес в научное издательство за последние годы, кроме того, движение от печатного к электронному обеспечению журнальных статей и крупнейший вызов традиционным бизнес-моделям издателей подписного доступа – все это способствовало росту обеспечения открытого доступа, при котором статьи становятся доступными для читателей без барьеров на подписку или оплату. Без технической возможности использования Интернета для дешевой доставки содержания движение за открытый доступ будет мертворожденным.

Поскольку с рецензированием возможно изменение степени открытости доступа, необходимо провести четкие различия. В частности, статьи открытого доступа могут предоставляться для чтения бесплатно, но такие статьи могут оставаться защищенными авторским правом и лицензионными ограничениями. Они предотвратят все формы передачи, репродуцирования и многоцелевого использования помимо тех, которые допускаются принципами «справедливого использования» или «справедливого соглашения» закона об авторском праве посредством предотвращения многоцелевого использования содержания для извлечения текста, для репродуцирования выводимых работ или для коммерческих целей без письменного разрешения владельца авторского права.

Номенклатура, используемая для характеристики различных типов открытого доступа, вносит путаницу и применяется по-разному. Мое понимание этой сферы основано на особенно полезных положениях в блогах двух авторов по этому предмету [12,13], которые четко различают две перпендикулярные оси классификации:

- «Место нахождения копии статьи открытого доступа, характеризуемое как зеленый в отличие от золотого: *зеленый открытый доступ* означает, что статья доступна бесплатно с другого сайта, чем журнальный сетевой сайт, например из институционального хранилища, тогда как *золотой откры-*

*тый доступ* означает, что статья доступна бесплатно с собственным журнальным сетевым сайтом.

- Тип открытого доступа характеризуется как бесплатный в отличие от свободного: *бесплатный открытый доступ* означает устранение как стоимостного барьера, давая право видеть статью; в то время как *свободный открытый доступ* означает устранение как стоимостного барьера, так и, по меньшей мере, ослабление барьеров, ограничивающих многоцелевое использование, давая право использовать статью.

Поскольку оба типа подразумевают «свободный» (вероятно, двусмысленное слово), то бесплатный открытый доступ равнозначен «свободный, как выпивший человек», тогда как свободный открытый доступ равнозначен «свободный, как в разговоре». Бесплатный открытый доступ, таким образом, является необходимым, но недостаточным условием для свободного открытого доступа.

Основные декларации по открытому доступу, которые связаны с научным издательством, – Будапештская инициатива открытого доступа 2002 г., Берлинская декларация по открытому доступу к знанию в естественных и гуманитарных науках 2003 г. и Предложение по издательству открытого доступа 2003 г. (г. Бетесда, шт. Мэриленд, США) – все определяют открытый доступ как прозаично «свободный», но многие определения открытого доступа издателей приравниваются только к бесплатному открытому доступу. Очевидность того, что это означает быть полностью открытым, выражается определением *открытый* в «Основах открытого знания»:

«Часть содержания или данных является открытой, если любой человек может свободно использовать, использовать для многих целей и распространять ее, подчиняясь самое большее только требованию авторства и производному произведению».

Статьи, как зеленого открытого доступа, так и золотого открытого доступа могут быть либо бесплатного открытого доступа, либо свободного открытого доступа. Свободный открытый доступ наиболее четко описывается использованием подробной лицензии, такой как *Creative Commons Attribution License*, которая четко указывает, какие права даны читателю/пользователю статьи, или использованием документа об отказе от своих прав, тогда статья помещается в общедоступный домен. Без такой четкой спецификации, разумно полагать, что любая статья открытого доступа является только статьей «бесплатного открытого доступа».

Очевидно, что трудно свести эти две перпендикулярные классификации вместе к одной оценочной шкале, необходимой для второй из пяти звезд, поэтому я принимаю консервативный подход, допускающий, что статьи открытого доступа являются только статьями бесплатного открытого доступа, как бы не утверждал кто-то другой.

#### 0 – Открытый доступ отсутствует

Статья публикуется в журнале с доступом по подписке; недоступна для тех, кто не имеет персональной или институциональной подписки. От автора обычно требуется передача авторского права издателю, а самоархивирование публикуемой статьи в институциональном хранилище или в другом месте не разрешается.

#### 1 – Самоархивирование с зеленым/бесплатным открытым доступом

Доступный по подписке журнал позволяет автору самостоятельно издать препринт, пострецензуемый «постпринт» или копию статьи в PDF-формате издателя в институциональном хранилище или в другом месте без периода запрета, позволяя третьим лицам иметь

возможность читать текст без оплаты, в то же время не разрешая ее многоцелевого бесплатного использования. Копия статьи на журнальном сетевом сайте остается доступной только для подписчиков.

## 2 – Финансируемый зеленый/бесплатный открытый доступ

В ответ на полномочия от спонсора научного проекта, описанного в статье, и внесение издателю платы финансирующим органом издатель разрешает хранить копию статьи в *PubMed Central*, позволяя третьим лицам иметь возможность читать текст бесплатно, хотя не допуская бесплатного многоцелевого ее использования. Копия статьи на журнальном сетевом сайте остается доступной только для подписчиков.

## 3 – Золотой/бесплатный открытый доступ за авторскую плату

В обмен на внесение платы автором (или учреждением автора) издателю журнал публикует статью на журнальном сетевом сайте, так что третьи лица могут читать статьи бесплатно, хотя они не могут иметь бесплатного многоцелевого использования содержания. Журналы, которые дают золотой/бесплатный доступ за авторскую плату, могут дополнительно позволить самоархивирование публикуемой статьи в институциональном хранилище или в другом месте в виде статьи с зеленым/бесплатным открытым доступом.

## 4 – Золотой/свободный доступ с авторской платой

В обмен на внесение платы автором (или учреждением автора) издателю журнал публикует статью на журнальном сетевом сайте по лицензии *Creative Commons* или подобной лицензии авторства, делая статью свободно доступной для чтения и для, по меньшей мере, некоторых форм многоцелевого использования, включая самоархивирование с условием, что авторство остается за издательской «версией записи».

Эти категории требуют дальнейших пояснительных комментариев.

### Открытый доступ отсутствует

Всего несколько журналов с доступом по подписке, большинство из которых в области биомедицинских наук, позволяют автору размещать публикуемую версию статьи в институциональном хранилище или персональном сетевом сайте (т.е. самоархивирование с зеленым/бесплатным открытым доступом) после периода запрета обычно от шести до двенадцати месяцев. Информация об этом дается на сайте *SHERPA/JoMEO*. Однако своевременный доступ к вновь публикуемой научной информации остается строго ограниченным кругом подписчиков журнала.

### Самоархивирование с зеленым/бесплатным открытым доступом

В области физики, математики и вычислительной техники использование хранилища препринтов *ArXiv* Корнеллского университета считается обычным. *ArXiv* является образцовым хранилищем, поскольку все его содержание (контент) делается доступным либо по лицензии *Creative Commons Attribution*, либо по лицензии *Creative Commons Attribution-Noncommercial-ShareAlike* или размещается в общедоступном домене в связи с декларацией *Creative Commons Public Domain Declaration*. Таким образом, очевидно, что контент *ArXiv* имеет зеленый/свободный открытый доступ.

Однако для большинства научных дисциплин, где нет культуры хранения препринтов в одном предметно-специфическом открытом архиве до подачи публикации в журнал, зеленый открытый доступ является скромным четвертым среди выборов открытого доступа. Это происходит по двум причинам: в первую очередь,

из-за трудности, которую потенциальные читатели испытывают при нахождении версий статей открытого доступа, рассеянных по институциональным хранилищам (хотя новые услуги перекрестного поиска, такие как *CORE*, улучшают эту ситуацию), а также из-за того, что лицензионные договоренности о многоцелевом использовании такого содержания обычно не ясны.

Рассмотрим, например, Оксфордский научный архив – институциональное хранилище Оксфордского университета. Архив обеспечивает полезное руководство по авторскому праву в помощь авторам, хранящим работы в этом архиве. В этом руководстве обсуждаются ограничения авторского права, которые издатели могут распространять на публикуемые работы, и упоминается возможность использования лицензии *Creative Commons*. Тем не менее, его руководящий принцип для читателей, касающийся содержания Оксфордского научного архива, состоит в следующем:

«Полный текст многих из этих публикаций свободно доступен для использования в соответствии с авторским правом и разрешениями для конечного пользователя».

Институциональное научное хранилище Саутгемптонского университета *Eprints Soton* поддерживает идентичное предложение, тогда как институциональное научное хранилище Кембриджского университета *Dspace@Cambridge* – более ограниченное поверхностное предложение:

«Авторское право и другие права интеллектуальной собственности имеются на этом сайте в разделе *Deposited Works* (хранящиеся работы) и в любой другой сопроводительной документации и метаданных. Пока нет других примечаний, хранящиеся в *Dspace@Cambridge* работы делаются свободными для доступа, печати и импортирования в некоммерческих научных целях или только для личного изучения. *Вы не можете* в дальнейшем копировать, размножать, публиковать, ...или как-то еще *использовать* хранящиеся работы в целом или частично, или любым другим образом, или в любых СМИ (средствах) без специального письменного разрешения от соответствующих обладателей прав на хранящиеся работы».

Метаданные о совсем недавно хранящейся в Оксфордском научном архиве научной статье *Knight et al.* (2011), *The puzzle of migrant labour shortage and rural labour surplus in China* в формате PDF, не содержат информации о статусе открытого доступа к этой статье, который, как можно предположить, должен быть только бесплатным открытым доступом. Это подтверждается доступом к этой оригинальной статье на журнальном сетевом сайте издательства Elsevier (*China Economic Review*. - 2011. - Vol. 22, No 4. – P.585-600. – doi:10.1016/j.chieco.2011.01.006), где есть связь с «Permissions and Reprints» («Разрешения и репринты», которая ведет вас к странице, названной «RightsLink Copyright Clearance Centre»). Здесь любой человек может подсчитать стоимость многоцелевого использования статьи для других целей, чем чтение для исследования или личного изучения. Использование 15 печатных копий статьи в качестве курсового материала для обучения в Оксфордском университете будет стоить 26,74 ф. ст., тогда как использование 15 печатных копий статьи в целях подготовки в коммерческой организации будет стоить 384,34 ф. ст.

Кто-то может полагать, что эта ситуация подобна ситуации в других институциональных хранилищах и что публикации доступны только зеленым/бесплатным открытым доступом, если не существует определенной лицензии на свободный открытый доступ. Поэтому, как заключает Питер Мюррей-Руст: «Из-за



неплатежа, если автор-самоархивист не приложит специального усилия, читатель (публикации в институциональном хранилище) не имеет прав на использование хранящейся публикации» [13].

#### *Финансируемый/зеленый/бесплатный открытый доступ*

Хотя внесенная финансирующим органом плата издателю за хранение копии статьи в *PubMed Central* является существенной (обычно 3-5 тыс. долл. за статью), важно понимать, что читатель получает только права бесплатного открытого доступа. В частности, содержание недоступно для извлечения текста. Из всего содержания *PubMed Central*, почти 2,3 млн статей, только около 10 % в нем составляет то, что называется подмассивом открытого доступа, который имеет некоторую форму лицензии на свободный открытый доступ, полученный в основном от издателей, которые сами придерживаются политики золотого/свободного открытого доступа. Он (подмассив) состоит из библиографических перечней этих статей, а не из полного корпуса *PubMed Central*, из которого мы создали массив *Open Citation Corpus*, включающий из 6,3 млн библиографических ссылок, отражающих почти 20 % всех биомедицинских статей, опубликованных в 1950-2010 гг., в том числе самые важные статьи в каждой дисциплине. Опубликованные по свободной лицензии Creative Commons и представленные в формате RDF, эти ссылки доступны для *тщательного изучения человеком* и для автоматического запрашивания через конечный пункт SPARQL, а весь корпус также может быть импортирован для многоцелевого использования.

#### *Золотой/бесплатный открытый доступ с авторской платой*

Вносимая автором плата за золотой/бесплатный открытый доступ обычно существенна – в пределах 500-3250 долл. за статью. Сетевой сайт SHERPA/RoMEO обеспечивает подробности. Поскольку мои пять звезд предназначены для работы на уровне статьи, а не на уровне журнала, я не делаю здесь различия между статьями, которые открыто доступны на индивидуальной основе и к которым имеется доступ через подписку на журнал (иногда называемый журналом «гибридного» открытого доступа), и статьями в журнале «по длинно» открытого доступа, в котором все статьи открыто доступны. Издательство Elsevier называет первый из них «спонсируемым доступом», а журнал, в котором все статьи открыто доступны, – «журналом с авторской платой». Однако на уровне отдельной статьи нет различия: автор вносит плату, и каждый может читать статью свободно на сетевом сайте издателя.

Проблема, скорее, состоит в том, имеют или не имеют читатели права многоцелевого использования статьи или им гарантирован только золотой/бесплатный открытый доступ. Например, политика издательства Elsevier, касающаяся статей открытого доступа с авторской платой, представлена на его странице под рубрикой Terms and Conditions (Термины и условия), которую каждый человек может достичь, «кликнув» эту сноску на домашней странице его единственного журнала с открытым доступом *International Journal of Surgery Case Reports*. Среди других ограничений это гласит:

«Все содержание, находящееся на этом сайте или доступное с него, ... является собственностью издательства Elsevier или его лицензиатов и защищено авторским правом, торговой маркой и другими законами интеллектуальной собственности и несправедливой конкуренции. Вы не можете копировать, представлять на экране, распространять ... или создавать

другие производные работы из ... всего или какой-либо части этого содержания ... за исключением специально разрешенного использования по условиям Terms and Conditions, соответствующей лицензии или соглашения о подписке, либо санкционированного нами доступа. Если у вас нет специального санкционированного нами доступа, вы не можете автоматически искать, стирать, извлекать, углублять связь или индекс какого-либо содержания».

Естественно, в обмен на 3 тыс. долл. за статью, выплаченные спонсором или автором статьи издательству Elsevier, читатели получают только золотой/бесплатный открытый доступ. Такая незначительная сумма, возможно, сказалась на факте, который имел место в 2009 г., когда в 450 журналах издательства Elsevier, предлагающих «спонсируемый доступ» к своим статьям, только 515 статей «спонсируемого доступа» были опубликованы.

#### *Золотой/свободный открытый доступ с авторской платой*

Авторская плата за золотой/свободный открытый доступ также обычно является существенной – опять в пределах 500-3250 долл. за статью, но в этом случае издатель позволяет третьим лицам как читать все его статьи на журнальном сетевом сайте бесплатно, так и использовать содержание для многих целей. Как указывает Питер Субер [12], свободный открытый доступ включает ряд возможностей, в соответствии с которыми гарантируется разрешение многоцелевого использования. Например, можно использовать содержание статьи для создания производной работы, но если она не будет использоваться в коммерческих целях. Рамки многоцелевого использования определяются характером лицензии, по которой статья публикуется. Это, таким образом, позволяет очень условно назвать свободный открытый доступ «полным открытым доступом».

Оба крупнейшие издатели журналов открытого доступа в области биомедицинских наук – *BioMed Central* и *The Public Library of Science (PLoS)* – используют наиболее позвольительные в отношении открытого доступа лицензии для всех работ, которые они публикуют. (Лицензионное соглашение The BioMed Central Open Access имеет непохожее название, но является другой идентичной лицензией Creative Commons, применяемой *PLoS*). По такой лицензии авторы оставляют за собой владение авторским правом на их содержание, но разрешают третьим лицам импортировать, многообразно использовать, перепечатывать, изменять, распространять и/или копировать содержание для любой цели, включая коммерческую цель, до тех пор, пока авторы оригинала и источник цитируются. Разрешения от авторов или издателей не требуется. Это, очевидно, является самой полезной ситуацией для различных потенциальных пользователей публикуемых статей. Мы смогли применить *семантические обогащения* к статье [14] *PLoS Neglected Tropical Diseases* [4], что стало возможным, потому что эта статья была опубликована по такой лицензии.

### **Обогащенное содержание**

**Использование полного потенциала сетевых технологий и сетевых стандартов для обеспечения взаимодействия и семантического обогащения содержания вашей электронной статьи.**

Сетевая технология может использоваться для обеспечения различных семантических обогащений



научных журнальных статей, ссылок на внешние информационные источники, релевантные текстовому содержанию, и для различных типов взаимодействия пользователя, как описывается в рубрике *Семантическое издательство* во Введении.

#### 0 – Обогащения отсутствуют

Статья публикуется в электронном режиме как документ в формате HTML или PDF без особенностей, кроме тех, которые должны быть присущи печатному изданию той же статьи.

#### 1 – Активные сетевые связи

Электронная статья включает функциональные сетевые связи с информацией и сетевыми сайтами, непосредственно релевантными этой статье, например с домашними страницами авторов, каталогами поставщиков, базами данных и цитируемыми статьями.

#### 2 – Семантическое обогащение текста

Ключевые термины и понятия в тексте идентифицируются и различаются, например с помощью установки «мышь» над временными рабочими окнами, обеспечивающими семантические определения, формулы, связи с входами в базы данных и т.п., выполняемые живыми сетевыми службами. Библиографические списки представляют собой печатание ссылок.

#### 3 – «Живое» содержание

Статьи содержат интерактивные рисунки, семантические линзы, показывающие числовые данные под графиками, временные рабочие окна, обеспечивающие отрывки из цитируемых статей, релевантных текстовым контекстам ссылок, упорядоченные библиографические списки и другие формы интерактивного содержания.

#### 4 – Слияния данных (« *mash-ups* »)

Данные в статье объединяются с уже существующей информацией (например, с подобными данными из других статей или баз данных), с географическими данными места, из которого статья доступна, как, например, файлы KML для визуализации в Google Maps и т.п.

Поскольку возможны разные типы семантического обогащения, и некоторые средства для их достижения были уже подробно описаны в [4,15], они не будут обсуждаться здесь далее. Как говорилось ранее, некоторые издатели и редакторы журналов предпринимают такие обогащения. Однако их можно лучше всего достичь во время написания статьи автором. Когда статья пишется, авторы могут легко достичь быстрых успехов в терминах функциональности путем обеспечения обильных связей с внешними сетевыми ресурсами (например, с их собственными домашними страницами, каталогами поставщиков реагентов и цитируемыми статьями). Плагин (независимо комплектуемый программный модуль) открытого источника в Word 2007 был опубликован, чтобы обеспечить семантическую разметку названных вещей в соответствии с избранными онтологиями [16], и можно надеяться, что другие такие семантические авторские средства скоро станут доступными.

### Доступные наборы данных

**Обеспечьте, чтобы все данные о результатах вашего отчета полностью публиковались по открытой лицензии с достаточными метаданными для возможной их повторной интерпретации и многоцелевого использования.**

В соответствии с Брюссельской декларацией по издательству в области науки, техники и медицины научные издатели строго придерживаются принципа, что научные

данные, связанные с журнальными статьями, должны быть сделаны свободно доступными, чтобы дать возможность проверять данные и обосновывать утверждения, сделанные в статье, и позволить многоцелевое использование данных в других контекстах. В частности, если исследование было проведено с публичным финансированием, теперь в значительной степени считается, что научные данные должны рассматриваться как общественный продукт [17, 18], и предлагаются механизмы для облегчения их публикации [11, 19, 20]. Однако в этом достойном похвалы энтузиазме за открытость важно признать личное время и усилия, затраченные исследователями, которые открыли или создали эти данные, и их моральное право иметь первый шанс изучать, публиковать и получать научную пользу из этих данных до публикации их для пользы других.

Важно также подчеркнуть, что термин «данные» должен здесь пониматься в очень широком смысле, включая любые результаты научного исследования из всего текста журнальной статьи и сверх него. Такие данные могут включать изображения, звукозаписи и видеозаписи, графики и диаграммы, анимации и модели, математические модели, протоколы и рабочий поток и программные средства, а также числовые наборы данных.

Принципы, как лучше сделать данные доступными в сети, уже были описаны Тимом Бернерсом-Ли в его работе о пяти звездах связанных открытых данных. Некоторое перекрытие неизбежно, но следующие рейтинги (оценки) предназначены отразить *природу* делания данных доступными, а также где, когда и для кого эта доступность гарантирована.

#### 0 – Нет опубликованных данных

Только те данные доступны, которые могут быть получены читателем из самой статьи. Рисунки и таблицы недоступны отдельно для импортирования и не являются какими-либо поддерживающими наборами данных.

#### 1 – Доступные дополнительные информационные файлы

Дополнительные информационные файлы доступны с журнального сетевого сайта, и/или рисунки и таблицы, содержащие научные данные в статье, доступны для импортирования. Однако форматы этих данных не приспособлены для многоцелевого использования, например рисунки и таблицы из журнальных статей PLoS доступны только в изобразительном формате TIFF или PNG.

#### 2 – Импортирование статейных данных в действующей форме

Данные, содержащиеся в рисунках и таблицах статьи и в ее дополнительных информационных файлах, доступны в соответствующих действующих форматах, например числовые данные в импортируемых числовых электронных таблицах и файлах CSV.

#### 3 – Основные публикуемые наборы данных

Полные научные наборы данных, созданные в ходе исследовательского проекта, из которых отбирается поднабор данных, включаемый в публикуемую статью, публикуются в постоянном архиве или хранилище с уникальным разрешенным идентификатором (например, URI или DataCite DOI) с лицензией на открыто доступные данные или отказом от права CC Zero и посвящением публичной сфере, а также с достаточными описательными метаданными для возможности их перепроверки и многоцелевого использования.

Вместе с журналом наборы данных, поддерживающие журнальную статью, делаются доступными для рецензентов, чтобы помочь им в оценке статьи. Это обычно делается конфиденциально до публикации этих наборов данных одновременно со статьей.

Очень важно то, где публикуются данные. Авторы должны иметь в виду очень неудовлетворительный характер журнальных файлов дополнительной информации в качестве хранилищ ценных научных данных в терминах открытости, поиска, хранения и надежной устойчивости [21-23]. В качестве безопасной гавани для публикуемых данных они взамен должны присмотреть институциональные хранилища или лучше предметно-специфические базы данных и хранилища. Например, хранилище Dryad Data Repository содержит биологические наборы данных, связанные с рецензируемыми журнальными статьями, что делает их доступными для предпубликационных рецензентов, а затем публикует их, либо в то же время, что и статью, либо после необязательного запретного периода, по лицензии Creative Commons CC Zero с отказом от права на открытые данные и с DataCite DOI, чтобы позволить надлежащее цитирование и заслужить научное доверие.

### Машиночитаемые метаданные

**Публикуйте машиночитаемые метаданные, описывающие как вашу статью, так и цитируемые вами ссылки, таким образом, чтобы они могли быть отыскиваемыми автоматически.**

Сегодня издатели применяют множество соответствующих информационных моделей, основанных на XML, и определений типов документов для разметки компонентов электронного документа (название, состав авторов, реферат и т.д.), чтобы помочь издательскому процессу, но все они, довольно часто даже эти основные метаданные, делаются недоступными читателям, которые получают только версию статьи в формате PDF.

Современные сетевые методы управления информацией, применяющие стандарты W3C, такие как RDF и OWL2, позволяют такой информации быть закодированной с использованием стандартных словарей, чтобы дать возможность компьютерам запрашивать метаданные и включать сетевую информацию из множества ресурсов в автоматическом режиме. Онтологии SPAR (Semantic Publishing and Referencing) – только некоторые из словарей, применяемых с такой целью, чтобы описывать научные публикации [9].

С использованием этих сетевых стандартов и словарей можно обеспечить семантические описания структурных и риторических компонентов статьи, применяя DoCO (Document Components Ontology), и создавать и публиковать машиночитаемые метаданные в формате RDF, которые описывают саму журнальную статью, т.е. которые кодируют стандартную библиографическую информацию, определяющую статью (авторы, год издания, название статьи, название журнала, номер тома, номера страниц, DOI и т.п.), применяя FaBiO (FRBR-aligned Bibliographic Ontology) и BiRO (Bibliographic Reference Ontology). Также подобным образом можно кодировать библиографическую информацию для всех ссылок в библиографическом списке статьи и использовать онтологию CITO (Citation Typing Ontology), чтобы подтвердить *существование* связи (ссылки) между цитирующей и цитируемой статьями (например, <Статья А> cito: цитирует <Статью В>), а также чтобы

описать тип или характер этой ссылки, как фактически, так и риторически [8,9].

Конечно, потребность в машиночитаемых метаданных не заканчивается на этом. Существует растущий ряд контрольных перечней и минимальных информационных стандартов, описывающих информацию, которая должна быть включена в научную публикацию, или определяющих метаданные для описания статей или наборов данных внутри определенных областей. Одним таким примером служит стандарт MPIDI (Minimal Information standard for reporting an Infections Disease Investigation). С использованием MPIDI Editor метаданные могут быть структурированы в соответствии с MPIDI, чтобы описать исследование инфекционной болезни и его научные результаты, включая журнальные статьи и научные наборы данных. Для первого случая метаданные могут включать предложения об основных гипотезах научного исследования и принципиальные выводы, описанные в статье, в добавление к обеспечению практических предложений, касающихся характера заболевания, числа пациентов и т.п.

Доступность метаданных статьи может оцениваться по следующей шкале:

#### 0 – Доступных метаданных нет

Статья публикуется только в виде документа в формате PDF. Разметка с помощью языка XML, применяемая издателями во время подготовки статьи, редактирования и производства рабочего потока, не публикуется.

#### 1 – Доступная структурная разметка

Разметка с помощью языка XML структуры документа, использующая определение типа документа (DTD) издателя, обозначающее название, состав авторов, реферат и т.п., включается в версию XHTML электронной статьи.

#### 2 – Доступные библиографические и ссылочные метаданные

Полные библиографические метаданные статьи и полные метаданные ссылок для ее библиографического списка публикуются как импортируемые машиночитаемые файлы или как связанные открытые данные в тройном хранении.

#### 3 – Богатая встроенная разметка

Дополнительная структурная, риторическая и семантическая разметка встроена в электронную статью, закодирована в формате RDF или в подобном машиночитаемом формате.

#### 4 – Структурированное резюме статьи

Машиночитаемое резюме ключевых фактов, гипотез, данных и выводов статьи делается свободно доступным как в человекочитаемой, так и машиночитаемой форме, основанной на минимальном информационном стандарте, соответствующем предметной области.

Существует несколько способов, посредством которых такие метаданные могут стать доступными. Как указывалось ранее, структурная разметка может включаться в сам документ XHTML. С использованием формата RDF можно встроить семантическую разметку в сетевой документ таким образом, что эти машиночитаемые метаданные становятся частью сети связанных открытых данных. Другие возможности встроенной разметки существуют с применением *микроданных* в документах HTML5. Альтернативно, библиографические и ссылочные метаданные могут сопровождать соответствующую журнальную статью как прилагаемые электронные файлы RDF: такие файлы сопровождают мою статью [9] и нашу улучшенную версию статьи [14]. Однако, как и для

научных наборов данных, связанных со статьей, это является преимуществом, если соответствующие файлы метаданных также принадлежат соответствующим хранилищам связанных открытых данных, таким как Open Bibliography Project и Open Citation Corpus.

Подробные метаданные, описывающие содержание статьи, могут создать основу для структурного цифрового резюме, описывающего суть статьи в обоих форматах - человекочитаемом и машиночитаемом, которые могут быть опубликованы в качестве открытого научного отчета в журнале с открытым доступом к данным (точнее в этом случае, «журнале метаданных»), тогда как отдельные фактические положения статьи могут издаваться как нанопубликации [24].

## ОЦЕНКА ПУБЛИКУЕМЫХ СТАТЕЙ ПО ПЯТИ ЗВЕЗДАМ

Поскольку критерии, применяемые для шкал оценки, представленные в разделе «Пять звезд электронных журнальных статей», являются несколько спорными, и поскольку рейтинг определенной статьи по каждой оси может включать элементы субъективных суждений, эти пять звезд электронных журнальных статей обеспечивают концептуальную структуру, по которой можно судить о степени, достигнутой любой статьей, или о падении ее ниже идеала, который должен быть полезен для авторов, редакторов и издателей, которые теперь должны спросить себя: «Как мои электронные журнальные статьи оцениваются по этим пяти звездам?»

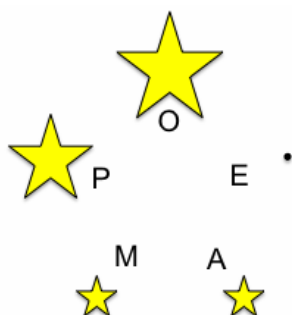
Как упражнение в «выпивании моего собственного шампанского» я оцениваю статью [14], какой она была до и после наших семантических обогащений, а также мои собственные недавние публикации, включая настоящую статью, чтобы обеспечить образцы. Каждая из них соотносится с каждой из пяти звезд электронных журнальных статей по пятизвездной шкале, данной в упомянутом выше разделе. Я представляю результаты как в виде диаграмм созвездия, в которых звезды имеют различные величины, так и в описательной форме с общим числовым рейтингом (оценкой) каждой статьи. (Полные библиографические элементы следующих статей приведены в разделе «Литература».)

### Reis et al. (2008) [14]

Журнал: *PLoS Neglected Tropical Diseases*. Издатель: Public Library of Science

Эта научная статья содержит различные типы анализируемых данных, касающихся факторов риска заболеть лептоспирозом для жителей городского населения в Сальвадоре и Бразилии. Основные непубликуемые сырые наборы данных содержат конфиденциальную информацию о здоровье, финансовом, семейном положении и занятии отдельных лиц.

Оценка: **Оригинальная версия:** <http://dx.doi.org/10.1371/journal.pntd.0000228>.



### Рецензия (P) 3

Ответная рецензия с возможностью для читателей делать постпубликационные замечания.

### Открытый доступ (O) 4

Золотой/свободный. Публикуемый в открытом доступе журнал с лицензией Creative Commons, обеспечил возможность внести семантические обогащения в оригинал авторов статьи [4].

### Обогащенное содержание (E) 0

Статья не имеет полезных сетевых связей в тексте и прямых связей с большинством цитируемых статей.

### Доступные наборы данных (A) 1

Рисунки и таблица в статье имеют свои DOI, но импортируются только как изображения, так что данные в графиках и таблице недоступны в действенной форме. Перевод реферата на португальском языке доступен как отдельно импортируемый документ в Word.

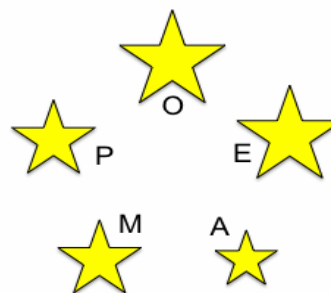
### Машиночитаемые метаданные (M) 1

Структурная разметка доступна в статье версии XHTML.

### Общая оценка: 9

Через неделю после публикации этой статьи [8] я взял ее для семантического улучшения, которое было предпринято при поддержке ее авторов и журнала *PLoS* и описано в [3,4]. Затем я перенес эту статью с ее улучшениями, чтобы она служила в качестве образца.

Оценка: **Семантически улучшенная версия:** <http://dx.doi.org/10.1371/journal.pntd.0000228.x001>.



### Рецензия (P) 3

Ответная рецензия на оригинальную статью с возможностью для читателей делать постпубликационные замечания.

### Открытый доступ (O) 4

Золотой/свободный. Семантически улучшенная версия, перенесенная с лицензией Creative Commons.

### Обогащенное содержание (E) 4

Многих типов, полностью описанных в [4].

### Доступные наборы данных (A) 2

Данные таблицы и двух рисунков, любезно предоставленные авторами и ставшие доступными в действенной форме как импортируемые в виде электронных таблиц Excel с их собственными DOI. Доступные различные дополнительные информационные файлы, включая приложения о слиянии данных (Data Fusion), в том числе Google Maps и связанные данные из других статей. Переведенный на португальский язык реферат доступен как отдельный документ в формате XHTML с семантической разметкой. Резюме статьи также доступно как отдельный документ. Однако положения, содержащиеся в резюме, не представлены в машиночитаемом формате и не имеют стандарта основных метаданных.

### Машиночитаемые метаданные (M) 3

Структурная и семантическая разметка в тексте XHTML; встроенная структура описания ресурса (RDF) обеспечивает основную библиографическую информацию; два импортируемых файла RDF с их собственными DOI сопровождают улучшенную статью, один из которых дает полные библиографические метаданные о статье, а другой обеспечивает библиографические элементы, ссылки на печатную информацию и частоту ссылок для всех цитируемых библиографических описаний.

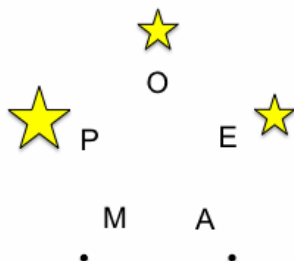
**Общая оценка: 16**

### Shotton [3]

Журнал: *Learned Publishing*. Издатель: Association of Learned and Professional Society Publishers

Эта статья описывает и анализирует состояние семантической публикации. В этой статье нет числовых данных.

Оценка:



Рецензия (P) 2

Ответная рецензия.

Открытый доступ (O) 1

Самоархивирование с зеленым/бесплатным доступом. Статья опубликована в журнале, изданном с доступом по подписке, что позволяет авторам публиковать постпринты в другом месте.

Обогащенное содержание (E) 1

Обильные сетевые связи в тексте и непосредственные связи со всеми цитируемыми статьями. Других форм семантического улучшения нет.

Доступные наборы данных (A) 0

Не применяются для этой статьи.

Машиночитаемые метаданные (M) 0

Отсутствуют. Статья доступна только в формате PDF.

**Общая оценка: 4**

### Shotton et. al. [4]

Журнал: *PLoS Computational Biology*. Издатель: Public Library of Science.

Эта статья описывает семантические улучшения, примененные к статье [8]. Как таковая она не содержит своих первичных научных данных.

Оценка:



Рецензия (P) 3

Ответная рецензия с возможностью для читателей делать постпубликационные замечания. (Редактор и один из рецензентов, которая высказала свое согласие с

авторами, дают исключительные подробности, чтобы помочь авторам улучшить статью, учитывая повторные круги рецензирования.)

Открытый доступ (O) 4

Золотой/свободный. Статья опубликована в журнале с открытым доступом по лицензии Creative Commons.

Обогащенное содержание (E) 1

Обильные сетевые связи в тексте, непосредственные связи со всеми цитируемыми статьями, рисунки имеют свои собственные DOI и связи с примерами семантического улучшения в нашей улучшенной версии статьи [8]. В самой статье семантические улучшения отсутствуют.

Доступные наборы данных (A) 1

Два дополнительных информационных файла, имеющие собственные DOI, опубликованы вместе со статьей, обеспечивая технические элементы семантических улучшений, сделанных в статье [8].

Машиночитаемые метаданные (M) 1

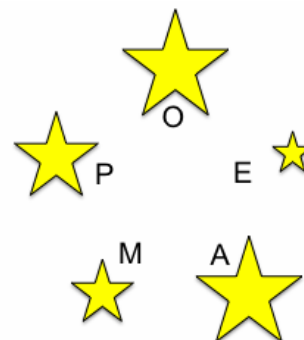
Структурная разметка доступна в версии статьи XHTML.

**Общая оценка: 10.**

### Shotton [8]

Журнал: *J. Biomedical Semantics*. Издатель: BioMed Central.

Оценка:



Рецензия (P) 3

Ответная рецензия с возможностью для читателей делать постпубликационные замечания.

Открытый доступ (O) 4

Золотой/свободный. Статья опубликована в журнале с открытым доступом по разрешительной авторской лицензии.

Обогащенное содержание (E) 1

Обильные сетевые связи в тексте, непосредственные связи со всеми цитируемыми статьями, рисунки и таблицы с их собственными DOI, но без семантической разметки текста и какой-либо интерактивности.

Доступные наборы данных (A) 4

Онтология CITO доступна в реальном времени с импортируемым человекомчитаемым дополнительным информационным файлом с его собственным DOI, обеспечивая дальнейшие пояснения; все наборы данных доступны для рецензентов.

Машиночитаемые метаданные (M) 2

Структурная разметка доступна в тексте XML. Импортируемые файлы RDF с их собственными DOI, один дает полные библиографические метаданные о статье, а другой обеспечивает библиографические элементы, ссылки на печатную информацию и частоту ссылок для цитируемых ссылок. Семантическая разметка, встроенный файл RDF и резюме статьи отсутствуют.

**Общая оценка: 14.**

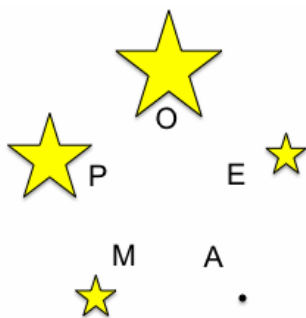
## Shotton (2012) [данная статья]

Журнал: *D-Lib Magazine*. Издатель: Corporation for National Research Initiatives (CNRI).

Эта статья является позиционной статьей, представляющей идею (в терминах *FaBIO ontology* – a fabio: proposition), и не содержит научных данных.

Несмотря на то, что журнал *D-Lib Magazine* не предусматривает формальной рецензии для своих статей и проводит политику непринятия к публикации статей, которые предварительно появились в другом месте, эта определенная статья вышла от существенных замечаний, сделанных коллегами (см. *Благодарность* для дальнейших подробностей) на *препринт этой статьи*, который был опубликован в *Nature Proceedings* для совещания корпорации Microsoft Research и Гарвардского университета в октябре 2010 г. под названием «Преобразование научной коммуникации». Я признателен редколлегии журнала *D-Lib Magazine* за ее гибкость в принятии этой статьи, несмотря на тот факт, что препринт уже был опубликован, поскольку полученные замечания стали эффективной постпубликационной ответной рецензией на содержание и качество окончательной статьи в *D-Lib Magazine*. В результате этих улучшений оценочные шкалы для некоторых из пяти звезд были изменены, обрекая оценки, данные ранее статьям Шоттона [2,4], быть заниженными относительно оценок, данных этим статьям в препринте.

Оценка:



Рецензия (P) 3

Постпубликационная ответная рецензия на препринт – см. комментарий выше.

Открытый доступ (O) 4

Золотой свободный открытый доступ без авторской платы.

Обогащенное содержание (E) 1

Обильные сетевые связи в тексте и ко всем ссылкам. Дополнительного семантического улучшения текста нет.

Доступные наборы данных (A) 0

Не применяются.

Машиночитаемые метаданные (M) 1

Только структурная разметка в HTML.

**Общая оценка: 9.**

Приведенные выше оценки показывают, что характер статьи будет влиять на полученную общую оценку. Например, обзоры и позиционные статьи без первичных научных данных всегда будут оцениваться низко с точки зрения доступных наборов данных.

Онтология пяти звезд электронных журнальных статей, доступная по адресу: <http://purl.org/spar/fivestars/>, является простой онтологией, написанной на языке OWL 2 DL, набором онтологий SPAR - Semantic Publishing and Referencing Ontologies (<http://purl.org/spar/>). Она предназначена для издателей и других лиц, желающих закодировать оценки пяти звезд, такие как описанные выше, в машиночи-

таемой форме с тем, чтобы они сопровождали другие машиночитаемые данные этой статьи. Следующий PDF-граф, приведенный в командной записи, дает оценки пяти звезд для этой статьи:

@prefix fivestars: <<http://purl.org/spar/fivestars/>> .  
@prefix xsd: <<http://www.w3.org/2001/XMLSchema#>> .

```
<http://dx.doi.org/10.1045/january2012-shotton>
  fivestars:hasPeerReviewRating
    "3"^^xsd:nonNegativeInteger ;
    fivestars:peerReviewRatingComment "Post-
publication responsive peer review of the pre-
print." ;
    fivestars:hasOpenAccessRating
    "4"^^xsd:nonNegativeInteger ;
    fivestars:openAccessRatingComment "Gold/libre
open access without author fee!" ;
    fivestars:hasEnhancedContentRating
    "1"^^xsd:nonNegativeInteger ;
    fivestars:enhancedContentRatingComment "Plen-
tiful web links in text and to all references. No
additional semantic enhancement of text." ;
    fivestars:hasAvailableDatasetsRating
    "0"^^xsd:nonNegativeInteger ;
    fivestars:hasMachine-readableMetadataRating
    "1"^^xsd:nonNegativeInteger ;
    fivestars:machine-
readableMetadataRatingComment "Structural markup
in HTML only." ;
    fivestars:hasOverallFiveStarsRating
    "9"^^xsd:nonNegativeInteger ;
    fivestars:overallFiveStarsRatingComment "The
nature of this article, being a position paper
rather than a research paper with primary re-
search data, has influenced the overall rating
obtained." .
```

Издательство Ubiquity Press указывает, что оно хочет применять такие оценки и давать каждой из своих публикуемых статей пятизвездочную оценку. Я предлагаю другим издателям поступать также.

*Благодарность.* Я очень признателен Бобу Дучарму, который, будучи воодушевленным пятью звездами связанных открытых данных Бернерса-Ли, предложил мне подойти с пятью звездами к семантической публикации после речи, озаглавленной “Applying XML and semantic technologies to liberate infectious disease data”, которую я произнес недавно в летней школе по XML в Оксфорде. Я благодарен Тане Грей и Кэтрин Флетчер за обратную связь после прочтения предварительного варианта этой статьи. Особую признательность выражаю Сильвио Перони, который настоял, чтобы я описал шкалы оценки для всех пяти звезд, и чьи предложения относительно рецензии и открытого доступа я включил; всем тем, кто участвовал в короткой, но живой дискуссии по препринту о пяти звездах по списку почтовой рассылки Beyond the PDF, особенно Камерону Нейлону, который предложил радикальный пересмотр моей оригинальной оценочной шкалы для рецензии, Филиппу Лорду за мудрые замечания, касающиеся постпубликационной рецензии и запроса комментариев, Питеру Мюррей-Русту за его настойчивость в том, чтобы тип лицензии, по которой публикации открытого доступа издавались, был решающим; и Брайану Хоулу из издательства Ubiquity Press за его комментарии и общий энтузиазм по поводу концепции пяти звезд. Их предложения составили эффективную постпубликационную рецензию *препринта* этой статьи, как упоминалось ранее, и я искренне благодарен им за время и усилия, потраченные ими, чтобы обеспечить эти ценные критические замечания.



## ЛІТЕРАТУРА

1. Berners-Lee T. Linked data. - <http://www.w3.org/DesignIssues/LinkedData.html>. - 2009.
2. Kurtz D., Parker G., Shotton D., Klyne G., Schroff F., Zisserman A., Wilkes Y. CLAROS — bringing classical art to a global public // Proc. IEEE e-Science Conference, Oxford, 9-11 December 2009. - P. 20-27. - <http://doi.ieeecomputersociety.org/10.1109/e-Science.2009.11>
3. Shotton D. Semantic publishing: The coming revolution in scientific journal publishing // Learned Publishing. - 2009. - Vol. 22. - P. 85-94. - <http://dx.doi.org/10.1087/2009202>. A post-print is available.
4. Shotton D., Portwin K., Klyne G., Miles A. Adventures in semantic publishing: Exemplar semantic enhancement of a research article // PLoS Computational Biology. - 2009. - Vol. 5. - P.e1000361. - <http://dx.doi.org/10.1371/journal.pcbi.1000361>.
5. Pafilis E., O'Donoghue S.I., Jensen L.J., Horn H., Kuhn M., Brown N.P., Schneider R. Reflect — augmented browsing for the life scientist // Nature Biotechnology. - 2009. - Vol. 27. - P. 508-510. - <http://dx.doi.org/10.1038/nbt0609-508>.
6. Wan S., Paris C., Dale R. Supporting browsing-specific information needs: Introducing the Citation-Sensitive In-Browser Summariser // Web Semantics: Science, Services and Agents on the World Wide Web. - 2010. - Vol. 8. - P.196-202. - <http://dx.doi.org/10.1016/j.websem.2010.03.002>.
7. Attwood T.K., Kell D.B., McDermott P., Marsh J., Pettifer S.R., Thorne D. Utopia documents: linking scholarly literature with research data // Bioinformatics. - 2010. - Vol. 26. - P. i568-i574. - <http://dx.doi.org/10.1093/bioinformatics/btq383>.
8. Shotton D. CiTO, the Citation Typing Ontology // J. Biomedical Semantics. - 2010. - Vol 1. (Suppl. 1). - P. S6. - <http://dx.doi.org/10.1186/2041-1480-1-S1-S6>.
9. Peroni S., Shotton D. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. (Submitted for publication). Preprint. - 2011.
10. de Waard A., Buckingham Shum S., Carusi A., Park J., Samwald M., Sándor Á. Hypotheses, evidence and relationships: The Hyper approach for representing scientific knowledge claims // Proceedings 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse (26 Oct 2009, Washington DC.) // Lecture Notes in Computer Science. - Springer Verlag: Berlin, 2009. - <http://oro.open.ac.uk/18563/>.
11. Bourne P., Clark T., Dale R., de Waard A., Herman I., Hory E., Shotton D. (on behalf Force 11 community). Force11 White Paper: Improving the Future of Research Communication and e-Scholarship. (Published 28 October 2011). - [http://force11.org/white\\_paper](http://force11.org/white_paper).
12. Suber P. Gratis and libre Open Access. SPARC // Open Access Newsletter (August 2008 issue). - <http://www.arl.org/sparc/publications/articles/gratisandlibre.shtml>.
13. Murray-Rust P. Green and gold open access? Libre and gratis. Reasons why readers and re-users matter. - Blog post.
14. Reis R.B., Ribeiro G.S., Felzenburgh R.D.M., Santana F.S.000, Mohr S., Melendez A.X.T.O., Queiroz A., Santos A.C., Ravines R.R., Tassinari W.S., Carvalho M.S., Reis M.G., Ko A.I. Impact of environment and social gradient on *Leptospira* infection in urban slums // PLoS Neglected Tropical Diseases. - 2008. - Vol. 2. - P. e228. - <http://dx.doi.org/10.1371/journal.pntd.0000228>.
15. Shotton D., Portwin K. Technical implementation of the semantic enhancements applied to Reis *et al.* // PLoS Neglected Tropical Diseases. - 2009. - Vol. 2, No.4. - P. e228. Supporting Information File S1 to Shotton *et al.* (2009). - <http://dx.doi.org/10.1371/journal.pntd.0000228.s009>.
16. Fink J.L., Fernicola P., Chandran R., Parastatidis S., Wade A., Naim O., Quinn G.B., Bourne P.E. Word add-in for ontology recognition: semantic enrichment of scientific literature // BMC Bioinformatics. - 2010. - Vol. 11. - P.103. - <http://dx.doi.org/10.1186/1471-2105-11-103>.
17. Boulton G., Rawlins M., Vallance P., Walport M. Science as a public enterprise: the case for open data // The Lancet. - Vol. 377. - P.1633-1635. - [http://dx.doi.org/10.1016/S0140-6736\(11\)60647-8](http://dx.doi.org/10.1016/S0140-6736(11)60647-8).
18. Wood J., Andersson T., Bachem A., Best C., Genova F., Lopez D.R., Los W., Marinucci M., Romary L., Van de Sompel H., Vigen J., Wittenburg P., Giaretta D., Hudson R.L. Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data; A submission to the European Commission, October 2010. - <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.
19. Greenberg J., White H.C., Carrier S., Scherle R. A meta-data best practice for a scientific data repository // Journal of Library Metadata. - 2009. - Vol. 9, No.3-4. - P.194-212. - <http://dx.doi.org/10.1080/19386380903405090>
20. Van der Graaf M., Waaijers L. A surfboard for riding the wave. Towards a four country action programme on research data. A knowledge exchange report. - <http://www.knowledgeexchange.info/Default.aspx?ID=469>.
21. Evangelou E., Trikalinos T.A., Ioannidis J.P. Unavailability of online supplementary scientific information from articles published in major journals // FASEB J. - 2005. - Vol. 19. - P.1943-1944. - <http://dx.doi.org/10.1096/fj.05-4784sf>.
22. Anderson N.R., Tarczy-Hornoch P., Bumgarner R.E. On the persistence of supplementary resources in biomedical publications // BMC Bioinformatics. - 2006. - Vol. 7. - P. 260. - <http://dx.doi.org/10.1186/1471-2105-7-260>.
23. Smit E. Abelard and Héloise: Why data and publications belong together // D-Lib Magazine. - 2011. - Vol. 17, No. 1/2. - <http://dx.doi.org/10.1045/january2011-smit>. (In particular, see Figure 12).
24. Groth P., Gibson A., Velterop J. The anatomy of a nanopublication // Journal of Information Services and Use. - 2010. - Vol. 30. - P. 51-56. - <http://dx.doi.org/10.3233/ISU-2010-0613>.

# Научное сотрудничество и расширение научной сети: измерение процессов глобализации в мире\*

**Роберт Я.В. ТЕЙССЕН**  
(Robert J.W. TIJSEN),

Центр научных и технологических исследований  
Лейденского университета,  
г. Лейден, Нидерланды

**Люддо ВАЛТМАН**  
(Ludo WALTMAN),

Центр развития регионального знания  
Лейденского университета, г. Гаага,  
Нидерланды

**Нес Ян ван ЭК**  
(Nees Jan van ECK),

Центр исследований по оценке, науке  
и технологии, Университет  
г. Стелленбос, Южная Африка

*В статье рассматриваются новая модель и аналитический инструмент для измерения и изучения современных процессов глобализации в науке сотрудничества - мире, в котором ученые, специалисты, техники и инженеры взаимодействуют в «сети» взаимосвязанных научных центров и сетях сотрудничества. Строительными блоками наших измерений являются города, где проводится научное исследование, как они указаны в адресах авторов научных публикаций. Единицей анализа служит географическое расстояние между этими городами. В нашем макроуровневом анализе тенденций, охватывающем 2000 - 2010 гг., мы наблюдаем, что расстояния научного сотрудничества увеличиваются, в то время как доля контактов сотрудничества с зарубежными городами выравнивается. Расстояния сотрудничества и темпы роста существенно различаются между странами и между областями науки. Применение измерений расстояния для сравнения и слежения за этими процессами открывает широкую дорогу для дальнейших изучений (как на мезо-, так и на микроуровне) того, как модели и тенденции научного сотрудничества двигают и формируют структуру связанности мировой науки.*

## ВВЕДЕНИЕ

Наука является неотъемлемой частью современного общества, которое вовлечено в процесс глобализации, быстро меняющуюся среду, в которой индивидуумы и организации постоянно взаимодействуют через территориальные границы и на больших расстояниях. Академическая наука, занятая базовыми исследованиями и осуществляемая независимыми учеными, всегда находилась на переднем крае этого эволюционного развития на протяжении двух последних веков. Производство и использование научного знания происходит через смешанные процессы творчества, коммуникации, сотрудничества и конкуренции. Одним из связующих элементов в этой структуре науки является совместное исследование, реализуемое индивидуумами и организациями в разных местах [1]. Одни сотрудничают без необходимости, другие из-за того, что ожидают дополнительную пользу от совместного использования ресурсов и вхождения в рабочую команду или просто из-за того, что это является требованием, установленным схемами научного финансирования. Небольшие организации с ограниченными ресурсами и оборудованием более склонны участвовать в пересекающей границы сотрудничестве, чем те, кто может полагаться на адекватную локальную научную базу или на-

циональную инфраструктуру. Географическое расстояние между научными партнерами является одной из определяющих основных характеристик этих взаимоотношений сотрудничества и сетей, которые могут возникать от неформальных небольших связей по сотрудничеству между близкими коллегами до формализованных и интенсивно использующих ресурсы крупномасштабных видов международного сотрудничества на большом оборудовании, таком как Международная космическая станция, Большой адронный коллайдер в Европейском центре ядерных исследований или Международный проект по термоядерным реакторам. С предполагаемым «исчезновением расстояния» [2] физическая близость должна стать менее определяющим фактором даже при наличии в настоящее время все больших возможностей сотрудничества в мировом масштабе при относительно низкой стоимости. Одновременно с этим эмпирические исследования продолжают показывать, что склонность искать партнеров по сотрудничеству поблизости, на месте или внутри страны, остается основным определяющим фактором в академической науке [3 - 6].

Модели и тенденции сотрудничества возникают из очень сложной адаптивной системы с миллионами, в значительной степени невидимых, взаимодействий и транзакций, больших неопределенностей относительно того, когда и как люди работают вместе, значительных эффектов случайных встреч (серендипность) и основных влияний внешних факторов, таких как механизмы финансирования, управленческие решения или поли-

\* Перевод Tijssen R.J.W., Waltman L., van Eck N.J. Research collaboration and the expanding science grid: Measuring globalization processes worldwide. – <http://arxiv.org/abs/1203.4194>. - 2012.

тические интервенции. Многочисленные взаимосвязанные определители находятся в действии, из которых только несколько главных структурных факторов могут быть смоделированы на макроуровне национальных научных систем: географические границы и язык, области науки, научная производительность и человеческие ресурсы, стадия экономического развития, международная ориентация и существование крупных осуществляющих научные исследования университетов, правительственных лабораторий или интенсивно занимающихся НИОКР компаний.

Исследования науки на макроуровне показывают, что среднее расстояние между сотрудничающими научными партнерами постепенно и постоянно увеличивалось в последние десятилетия [7, 8]. В этом смысле можно утверждать, что наука находится в процессе глобализации с точки зрения соглашений по транснациональному научному сотрудничеству, совместно используемых ресурсов, объединенной деятельности, миграции людей и потоков и обмена знаниями и навыками. Эти процессы переходов и интеграции происходят в мировой эволюционной системе, которая содержит мириады (несметное число) информационных, инфраструктурных, политических, культурных, социальных и экономических факторов, влияющих на структуры, модели и тенденции сотрудничества в науке.

Одним из ключевых определителей глобализации в науке является наличие потенциальных партнеров и желание сотрудничать в совместной научной деятельности. С открытием новых регионов и стран в мировой экономике и притоком новых участников в мировую науку, особенно из экономически возражающихся стран, вероятность нахождения новых возможностей для сотрудничества на длинном расстоянии значительно увеличилась в последние годы. Проводя параллель с крупномасштабными взаимосвязанными техническими инфраструктурами, подобными электрическим силовым сетям, можно сказать, что эта воображаемая структура возможных связей научного сотрудничества, эта «сеть науки» расширяется.

Определяя географические объекты как физические узлы в этой сети науки, мы можем ввести систему измерений расстояний между этими объектами для систематического изучения структурных и временных свойств связей сотрудничества. Основанная на расстоянии аналитическая модель охватывает спектр объектов: от локального (внутри города) сотрудничества до «глобального» сотрудничества на большом расстоянии, с «региональными» и «национальными» (внутри страны) видами сотрудничества в качестве промежуточных категорий. Использование городов в качестве единицы анализа также вводит опцию совокупности, особенно для регионов внутри страны, стран или даже целых континентов. Эти совокупные единицы можно выявить и сравнить в терминах их взаимной связности и степени глобализации.

Цель данного исследования – представить широкие модели и общие тенденции в мире. Мы понимаем, что эти тенденции и модели в значительной степени определяются показателями, специфичными для страны, или в терминах научных систем стран или внешних факторов, таких как геополитические или экономические достижения. Следовательно, наш анализ включает весь мир в целом в качестве совокупности, а также ограниченное число стран, рассматриваемых в качестве основных вкладчиков в мировую науку: Австралия, Канада, Франция, Германия, Япония, Нидерланды, Южная Корея,

Великобритания и США, а также группа стран БРИКС (Бразилия, Россия, Индия, Китай и Южная Африка). Исходя из нашей предыдущей работы [8], мы также изучаем тенденции глобализации на уровне широких областей науки.

В данной статье будут рассматриваться следующие общие вопросы:

1. Насколько фактический темп роста расстояний сотрудничества сравним с темпом роста в гипотетически полностью глобализованном мире?

2. Отличаются ли общие тенденции в расстояниях сотрудничества на уровне стран от таковых в расстояниях международного сотрудничества?

3. До какой степени глобализация в науке является феноменом, зависящим от области или от страны?

Содержание данной статьи построено следующим образом. Сначала мы приводим наш источник данных и методологию. Затем представляем наши эмпирические результаты и их интерпретацию в отношении исследуемых вопросов. Наконец, мы даем комментарии по проблемам дальнейшего исследования.

## ИСТОЧНИК ДАННЫХ И МЕТОДОЛОГИЯ

Модели и тенденции научного сотрудничества задокументированы сотнями тысяч научных публикаций, которые ежегодно появляются в открытой научной литературе. Значительное большинство этих публикаций содержит адреса мест работы авторов, так что сотрудничество прослеживается через многочисленные адреса авторов в совместно написанных авторами работах. Наша модель измерения основана на населенных пунктах, городках и крупных городах, упомянутых в этих адресах, которые совокупно будут с данного момента называться «городами». Принимая эту определенную географическую перспективу и уровень совокупности, город может рассматриваться как производитель научных публикаций, особый объект в ландшафте науки.

Нашим источником данных служит база данных Web of Science (WoS) издательства Thomson Reuter. Мы принимаем в учет все 11,1 млн представленных в WoS публикаций, появившихся в 2000 – 2010 гг., которые являются документом типа «статья» или «обзор» и которые, по меньшей мере, имеют один адрес в их перечне адресов. Для каждой публикации мы не учитывали адреса, которые многократно встречаются в перечне адресов. Важно отметить, что мы используем термин «адрес» для сочетания города и страны (а в случае США и Канады включается еще и штат) и игнорируем другие элементы адреса, такие как название организации, улица или почтовый индекс. Всего мы получили 18,8 млн адресов. Для каждого из этих адресов мы попытались определить соответствующие географические координаты (т.е. широту и долготу). Подхода геокодирования, которому мы следовали, является идентичным подходу, примененному нами в нашей более ранней работе [8]. Используя данный подход, мы смогли определить географические координаты 98,3 % всех адресов. Адреса, для которых географические координаты не могли быть определены, исключались из вычислений географических расстояний, обсуждаемых ниже.

Будучи основанными на описанных выше данных, вычисления были сделаны как для ряда выбранных стран, так и для мира в целом. Были вычислены пять следующих критериев:



- *Доля написанных в сотрудничестве публикаций.* Доля публикаций, в которых участвует страна, т.е. имеющих характер совместного написания. Публикация является совместной, если она имеет, по меньшей мере, два адреса.

- *Доля связей международного сотрудничества.* Доля связей сотрудничества, в которых участвует страна, т.е. международного сотрудничества.

- *Среднее географическое расстояние сотрудничества всех связей сотрудничества (MGCD All).* Среднее географическое расстояние между сотрудничающими городами вычислялось по всем связям сотрудничества, в которых участвует страна.

- *Среднее географическое расстояние сотрудничества только национальных (внутренних) связей сотрудничества (MGCD Domestic).* Среднее географическое расстояние между сотрудничающими городами вычислялось по всем национальным связям сотрудничества, в которых участвует страна. Следовательно, два города должны находиться в одной и той же стране.

- *Среднее географическое расстояние сотрудничества только международных связей сотрудничества (MGCD International).* Среднее географическое расстояние между сотрудничающими городами вычислялось по всем международным связям сотрудничества, в которых участвует страна. Следовательно, два города должны находиться в разных странах.

Отметим, что среднее географическое расстояние всех связей сотрудничества (MGCD All) равняется доле числа раз национальных связей сотрудничества (MGCD Domestic) плюс доля числа раз международных связей сотрудничества (MGCD International). Следовательно, MGCD All может рассматриваться как взвешенное среднее MGCD Domestic и MGCD International, где взвешивание определяется долей национальных и международных связей сотрудничества страны.

Подход фракционного подсчета применялся в вычислении вышеупомянутых критериев как на уровне публикаций, так и на уровне связей сотрудничества. Следующий пример иллюстрирует это. Предположим, что у нас есть три публикации, *X*, *Y* и *Z*. Публикация *X* имеет два адреса из страны *A*, один из страны *B* и один из страны *C*; публикация *Y* имеет три адреса из страны *A* и не имеет никаких других адресов; публикация *Z* имеет один адрес из страны *A* и у нее нет никаких других адресов. В нашем фракционном подходе подсчета публикация *X* приписывается стране *A* с весом  $\frac{1}{2}$  и странам *B* и *C* с весом  $\frac{1}{4}$  каждой. Публикация *X* включает  $4 \times (4 - 1) / 2 = 6$  связей сотрудничества. Каждая из этих связей имеет вес в  $1/6$ . Здесь есть одна национальная связь сотрудничества (между двумя адресами в стране *A*) и есть 5 международных связей сотрудничества. Страна *A* участвует в четырех из пяти международных связях сотрудничества. Публикация *Y* полностью приписывается стране *A*. Эта публикация включает  $3 \times (3 - 1) / 2 = 3$  связи сотрудничества, все они национальные. Каждая из этих связей имеет вес в  $1/3$ . Как и публикация *Y*, публикация *Z* полностью приписывается стране *A*, но в отличие от публикации *Y* публикация *Z* не включает никакого сотрудничества. Если публикации *X*, *Y* и *Z* являются единственными публикациями из страны *A*, тогда доля совместных публикаций страны *A* равняется  $(1/2 + 1) / (1/2 + 1 + 1) = 3,5$ . Доля международных связей сотрудничества страны *A* тогда равняется  $(4 \times 1/6) / (5 \times 1/6 + 3 \times 1/3) = 4/11$ .

Чтобы исследовать, до какой степени наука глобализована, обсуждаемые выше вычисления могут также производиться для гипотетически полностью глобализованного мира. Тогда реальный мир науки можно сравнить с гипотетическим миром. В настоящей статье мы рассматриваем гипотетический мир как основанный на простой произвольной модели сотрудничества. В этой модели сделаны два существенных предположения: 1) повсюду в мире исследователи, вероятно, в равной степени должны вовлекаться в сотрудничество и 2) партнеры по сотрудничеству выбирают друг друга произвольно, без обращения внимания на географическое расстояние между ними\*. Наша модель произвольного сотрудничества может быть проиллюстрирована следующим образом. Предположим, что в реальном мире мы имеем три адреса, *A*, *B* и *C*, и что эти три адреса дали 10, 20 и 60 публикаций, соответственно. Основываясь на предположениях нашей модели произвольного сотрудничества, можно ожидать, что здесь будет в три раза больше связей сотрудничества между адресами *A* и *C*, чем между адресами *A* и *B*. Также можно ожидать, что здесь будет в шесть раз больше связей сотрудничества между адресами *B* и *C*, чем между адресами *A* и *B*. Это означает, что из всех связей сотрудничества в нашей модели произвольного сотрудничества 10 % ожидается между *A* и *B*, 30 % - между *A* и *C* и 60 % - между *B* и *C*. Обсуждавшиеся выше вычисления для реального мира могут также производиться для гипотетически полностью глобализованного мира, основанного на нашей модели произвольного сотрудничества. Сравнение результатов, полученных для реального мира, с таковыми для гипотетического мира дает показатель степени, до которого может быть глобализован реальный мир.

При обсуждении наших эмпирических результатов в следующем разделе мы используем термин «произвольное географическое расстояние сотрудничества» (RGCD – random geographical collaboration distance) для рассмотрения среднего географического расстояния сотрудничества, вычисленного на основе нашей модели произвольного сотрудничества.

## ЭМПИРИЧЕСКИЕ РЕЗУЛЬТАТЫ

### Общие тенденции

Научное сотрудничество все в большей степени становится обычным в мировой науке. Доля написанных в сотрудничестве публикаций возросла с 35,0 % до 45,6 % за 2000 – 2010 гг. Однако доля международных связей сотрудничества, включая города в двух разных странах, осталась довольно стабильной за эти же годы, колеблясь в пределах 40,8 % – 41,8 %. Иначе говоря, научное сотрудничество между городами все еще растет, но доля международного сотрудничества выравнивается. Основываясь только на этом наблюдении, можно сделать поспешное заключение, что глобализация пришла к остановке. Но это игнорировало бы тот факт, что географическое расстояние между совместно публикующимися научными партнерами значительно увеличилось, как показано на рис. 1. MGCD All возросло на 5,2 % за 2000 – 2010 гг. - с 2 806 до 2 951 км.

\* Наша модель произвольного сотрудничества основана на подобных идеях, таких как критерий географической дисперсии (рассеивания), который мы использовали в нашем более раннем исследовании [8]. Схожие идеи также использованы в работе [3].

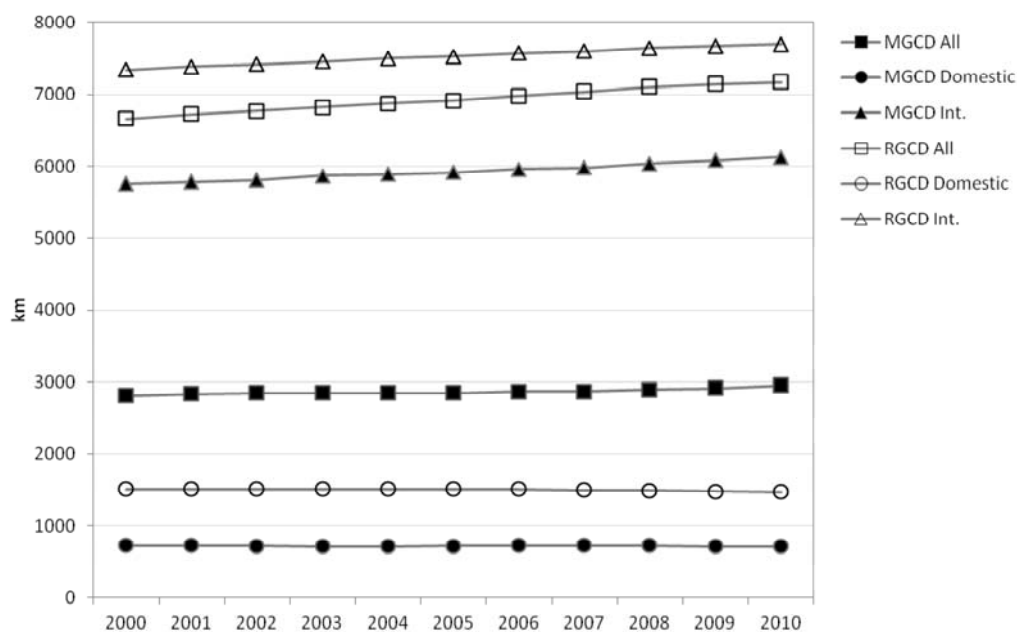


Рис. 1. Тенденции фактически существующих расстояний сотрудничества (MGCD) и расстояний произвольного сотрудничества (RGCD)

Если бы города должны были распространять свои связи сотрудничества произвольно через города-партнеры, активные в научном исследовании, тогда MGCD All было бы равно RGCD All, которое достигло 7 171 км в 2010 г. RGCD All увеличилось на 7,7 % в последнее десятилетие, таким образом опережая фактически существующие расстояния сотрудничества. Очевидно, что мы находимся на стадии роста процесса, где города либо постоянно вовлечены в сотрудничество на длинном расстоянии с существующими партнерами (усиливая установленное сотрудничество на длинном расстоянии), либо обнаруживая связи сотрудничества с новыми городами в глобальной научной сети, которые располагаются на более длинном расстоянии по сравнению с предыдущим средним расстоянием сотрудничества (новые длинные расстояния сотрудничества).

На рис. 1 MGCD и RGCD разделены в соответствии с типом сотрудничества - либо национальное, либо международное. Концентрируясь на международных партнерах, RGCD International поднялось до 7 696 км за 2000 – 2010 гг. с ростом в 4,8 %. MGCD International выросло немного больше, на 6,6 % до 6 126 км, ключ к тому, что действующие процессы глобализации выигрывают от расширяющейся научной сети, от новых возможностей для партнерства с городами на более длинном расстоянии. Наоборот, на национальном уровне MGCD Domestic и RGCD Domestic оба показывают снижение примерно на 2 %, которое предполагает тенденцию в странах вступать в сотрудничество с городами-партнерами, находящимися поближе.

В век глобально взаимосвязанной науки, где сотрудничество, совместно используемые научные методы и цифровая коммуникация становятся быстро растущей банальностью, можно рассчитывать увидеть конвергенцию (сходимость) MGCD и RGCD. Рис. 1 показывает нам иную картину. В 2010 г. RGCD All было в 2,43 раза больше, чем MGCD All. Следовательно, современные модели сотрудничества далеки от истощения возможностей сотрудничества в соответствии с нашей моделью произвольного сотрудничества. Модели национального сотрудничества обнаруживают схожий пробел между

MGCD и RGCD. RGCD Domestic в 2,06 раза больше, чем MGCD Domestic. В случае международного сотрудничества мы нашли, что MGCD и RGCD более сравнимые, отмечая, что сотрудничество через границы является относительно нечувствительным к географическому расстоянию. Тенденции на рис. 1 также показывают, что пробелы между MGCD и RGCD были постоянными на протяжении последнего десятилетия.

### Тенденции по областям науки

Общие тенденции и модели, отраженные на рис. 1, скрывают хорошо известные различия между областями науки. Каждая область характеризуется своим собственным исторически определенным набором параметров производства знаний, практик коммуникации и систем управления, которые совместно определяют форму, размер и динамику моделей научного сотрудничества. Например, благодаря использованию совместных научных инструментов, астрономы по необходимости демонстрируют большую склонность к сотрудничеству на длинном расстоянии, чем, скажем, инженеры-строители или социологи, как показано в нашем предыдущем исследовании [8].

До какой степени области науки являются решающим параметром в наших критериях глобализации? В таблице приводится общая статистика для мировой науки (ALL), а также для четырех широких областей науки: естественные науки (NAT – natural sciences), науки о жизни, медицина и здравоохранение (LIFE – life, medical and health sciences), инженерия, вычислительная техника и математика (ENG – engineering science, computer science and mathematics) и общественные и поведенческие науки (SOC – social and behavioral sciences)\*. Вместе

\* Эти четыре широкие области были созданы путем группирования вместе предметных категорий WoS на основе моделей цитирования. Это было сделано с использованием программного обеспечения VOSviewer ([www.vosviewer.com](http://www.vosviewer.com)). Окончательное отнесение предметных категорий к четырем широким областям было сделано полуавтоматически. Основная информация о сфере разных категорий была также принята во внимание.

эти четыре области охватывают всю базу данных WoS, за исключением искусства и гуманитарных наук и многодисциплинарных журналов, таких как *Nature* и *Science*. С MGCD All в 3 236 км ENG в среднем включает сотрудничество на самых длинных расстояниях. LIFE имеет самое короткое расстояние сотрудничества, с MGCD All на 14 % ниже, чем ENG. Однако заметим, что доля совместно написанных публикаций значительно выше в LIFE, чем в ENG (48 % и 43 %, соответственно). Кроме того, LIFE лидирует в условиях RGCD All, указывая, что по сравнению с другими широкими областями научные исследования LIFE являются самыми географически рассеянными.

Что касается общих тенденций по областям, то все области показывают рост и в MGCD All и в RGCD All. Область SOC пока растет быстрее всех: MGCD All поднялось на 16 %, а RGCD All - на 17 %. Интересно, что MGCD International снизилось на 2 %, а MGCD Domestic - на 8 %. Следовательно, рост в MGCD All происходит полностью благодаря большому увеличению (на 32 %) в доле международных связей сотрудничества. NAT и LIFE - обе области показывают значительные темпы общего роста. ENG слегка отстает (задерживается) в 2000 – 2010 гг., всего лишь с небольшим процентом роста, который может означать эффект выравнивания. Что касается доли совместно написанных публикаций и доли международных связей сотрудничества, то мы обнаружили, что NAT отстает от других областей в терминах низких темпов роста, которые опять можно отнести за счет возникновения процессов выравнивания.

## Тенденции по странам

Модели и процессы сотрудничества, вероятно, должны быть до определенной степени специфичными для стран. Небольшие и менее развитые страны могут быть, более вероятно, заняты в сотрудничестве через границы, чем большие продвинутые страны. Географически изолированные страны вынуждены сотрудничать через более длинные расстояния, чем страны, которые хорошо расположены на континентах или территориях, интенсивно занимающихся наукой. Более полное понимание процессов глобализации во всем мире требует проникновения в модели и тенденции сотрудничества отдельных стран. Наш сравнительный анализ охватывает выборку из 14 стран, которые все являются членами Организации экономического сотрудничества (OECD) или имеют партнерские отношения с OECD. Эта выборка представляет ряд ведущих мировых экономик и многие из самых крупных стран в сфере науки.

Рис. 2 показывает долю публикаций, написанных в сотрудничестве, и долю международных связей сотрудничества для каждой из 14 стран, где добавлено среднее

мировое значение в качестве общего контрольного значения. Существует четкое деление на две различные подгруппы в соответствии с долей международных связей сотрудничества в стране, т.е. связей сотрудничества, включающих научного партнера в зарубежном городе: 20-35 % в странах «национальной ориентации» и 45-60 % в странах «международной ориентации». Рис. 3 добавляет соответствующие данные о тенденциях. Несколько стран демонстрируют снижающуюся долю международных связей сотрудничества. Большинство из этих отклонений происходит в группе стран с национальной ориентацией. Бразилия представляет собой крайний случай с очень значительным снижением в доле международных связей сотрудничества.

На рис. 4 представлены MGCD Domestic и MGCD International для каждой страны в 2010 г. Различия между странами в терминах географических расстояний являются значительными. В случае Нидерландов национальные партнеры находятся в среднем на расстоянии в пределах 100 км друг от друга, тогда как международные партнеры в среднем находятся на расстоянии в 3 500 км. В другом крайнем случае австралийские ученые имеют национальные и международные расстояния сотрудничества в 1 000 и 13 000 км, соответственно. Научные системы Канады, Китая, России и США также характеризуются длинными расстояниями между национальными партнерами, с средними значениями более 800 км.

На рис. 5 показаны соответствующие темпы роста за 2000 - 2010 гг. в странах, которые распределены по следующим четырем квадратам:

- Национальное и международное увеличение расстояний сотрудничества (Канада и Франция представляют наиболее крайние случаи в этом квадрате).
- Национальное уменьшение и международное увеличение (Китай).
- Национальное увеличение и международное уменьшение (Австралия, Япония и Южная Африка).
- Национальное и международное уменьшение (Индия).

Растущие национальные и международные расстояния сотрудничества Франции показывают, что научное партнерство происходит на значительно длинных расстояниях. Хотя рост MGCD Domestic составляет только 20 км, это может быть сигналом того, что Франция немного отходит от своей традиционной научной системы, сосредотачивающейся в Париже. В случае Китая MGCD International увеличилось значительно, а MGCD Domestic снизилось. Эта тенденция отражает одновременный процесс увеличения национального сотрудничества с партнерами, находящимися на относительно близком расстоянии, и вступления в зарубежное партнерство на более длинных расстояниях [9].

**Тенденции в фактических расстояниях сотрудничества (MGCD) и расстояниях произвольного сотрудничества (RGCD): расстояния в 2010 г. (в км) и соответствующие изменения в 2000 – 2010 гг. (в интервалах, %)**

|                                           | ALL       | ENG        | LIFE      | NAT       | SOC       |
|-------------------------------------------|-----------|------------|-----------|-----------|-----------|
| MGCD All                                  | 2951 (5)  | 3236 (2)   | 2768 (6)  | 3032 (5)  | 3034 (16) |
| RGCD All                                  | 7171 (8)  | 7008 (3)   | 7168 (10) | 6981 (4)  | 6546 (17) |
| MGCD Domestic                             | 715 (-2)  | 680 (-8)   | 699 (0)   | 680 (0)   | 933 (-8)  |
| RGCD Domestic                             | 1474 (-2) | 1304 (-13) | 1528 (1)  | 1271 (-3) | 1588 (-3) |
| MGCD Int.                                 | 6126 (7)  | 6264 (2)   | 6224 (6)  | 5862 (10) | 6485 (-2) |
| RGCD Int.                                 | 7696 (5)  | 7421 (1)   | 7774 (6)  | 7433 (4)  | 7604 (3)  |
| Написанные в сотрудничестве публикации, % | 46 (30)   | 43 (25)    | 48 (38)   | 47 (20)   | 41 (39)   |
| Международные связи сотрудничества, %     | 41 (0)    | 46 (2)     | 37 (2)    | 45 (-4)   | 38 (32)   |

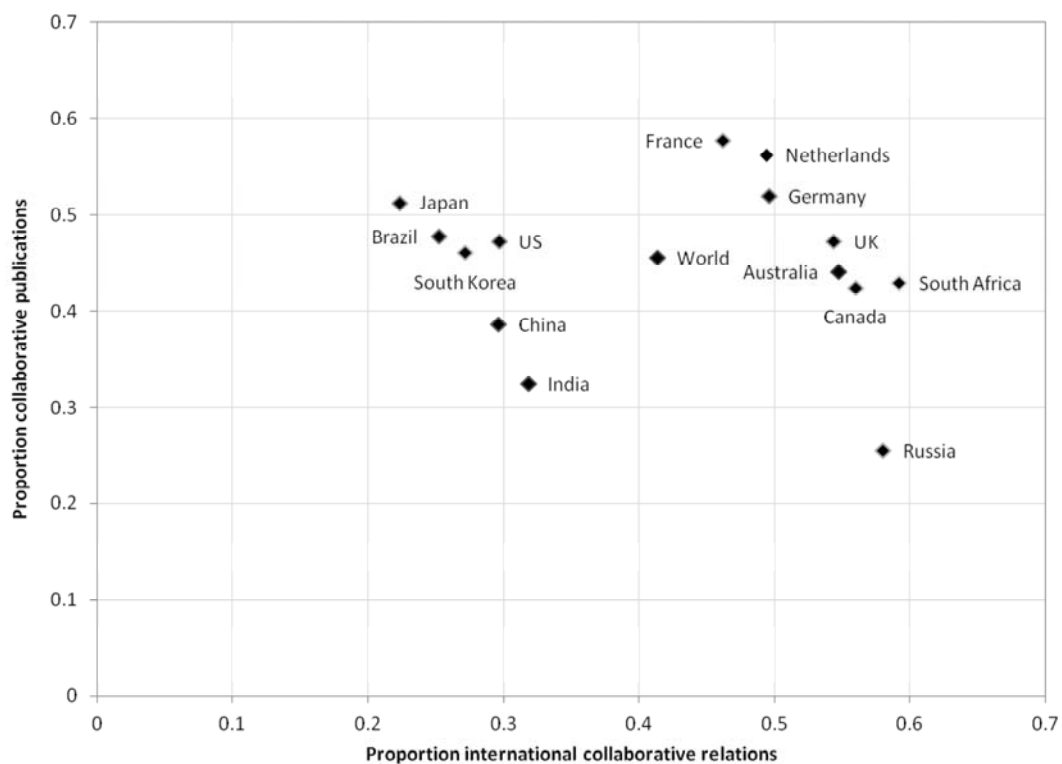


Рис. 2. Доля публикаций, написанных в сотрудничестве, и доля международных связей сотрудничества для 14 выбранных стран. Данные основаны на публикациях за 2010 г.

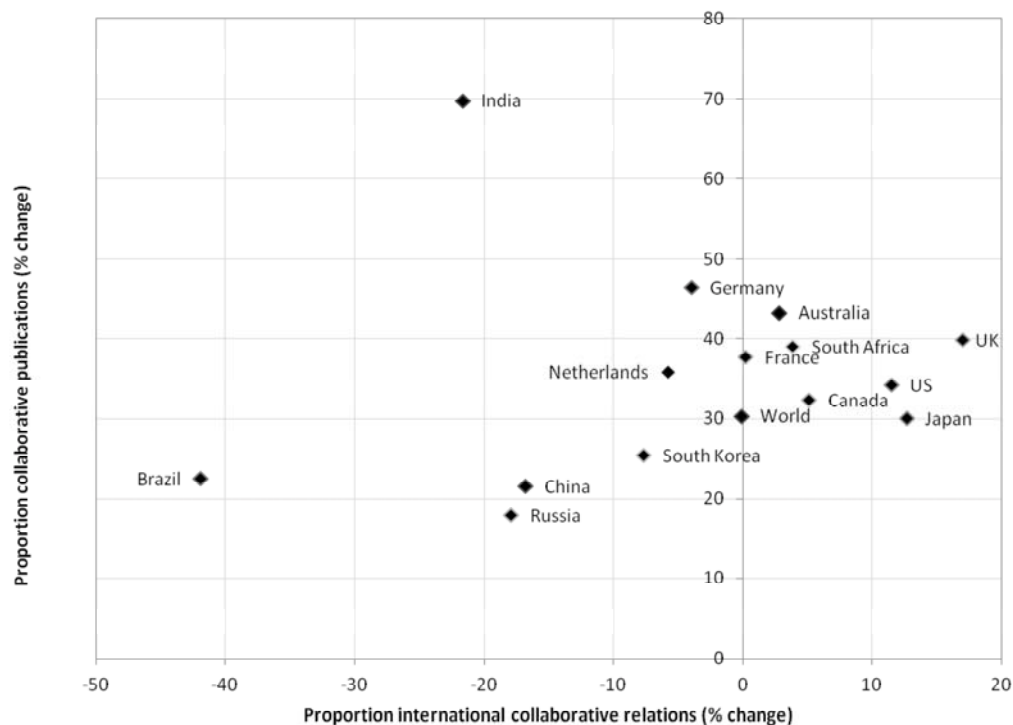


Рис.3. Соответствующее изменение за 2000-2010 гг. в доле публикаций, написанных в сотрудничестве, и доля международных связей сотрудничества для 14 выбранных стран

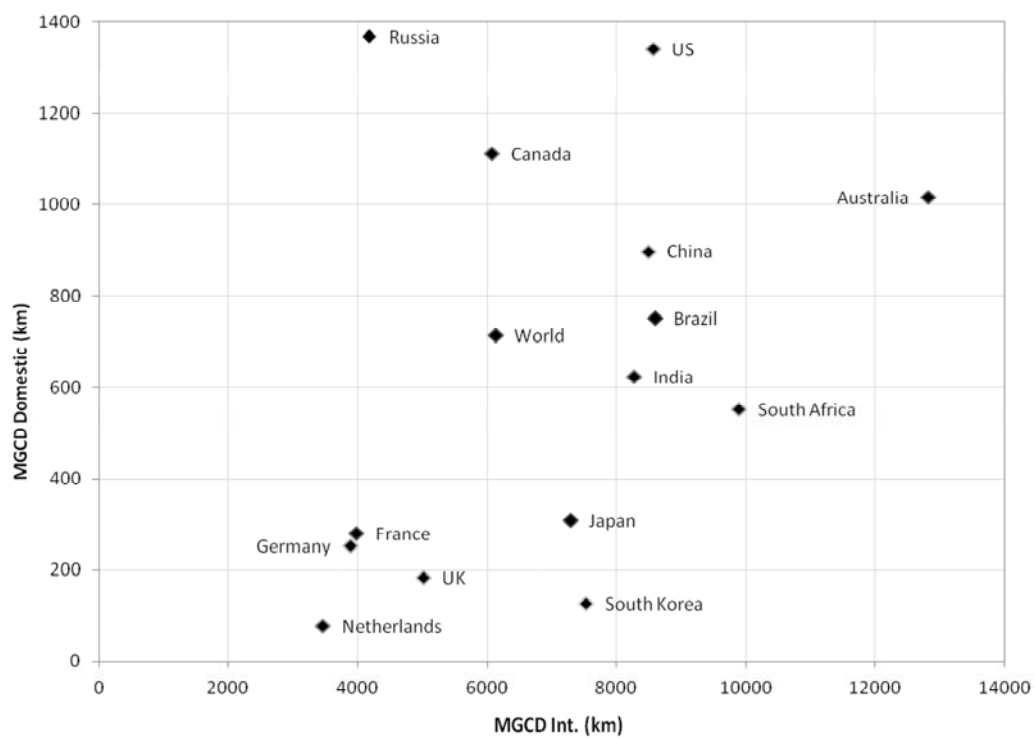


Рис.4. MGCD Domestic и MGCD International для 14 выбранных стран.  
Данные основаны на публикациях за 2010 г.

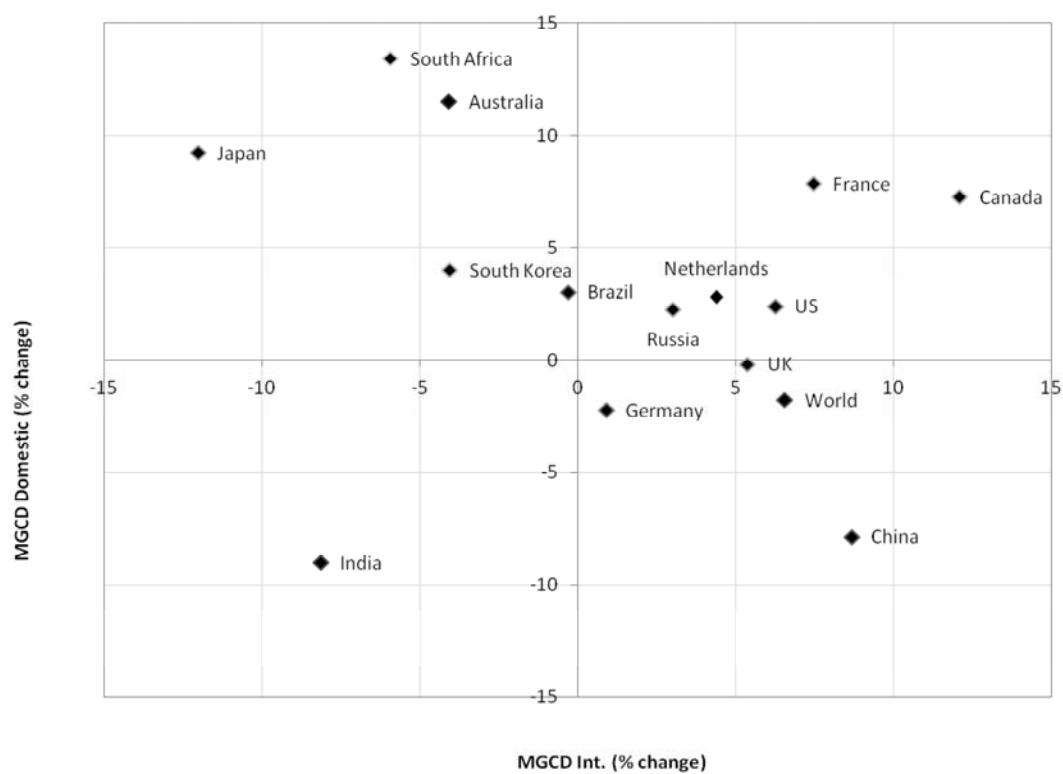


Рис.5. Соответствующее изменение за 2000-2010 гг.  
в MGCD Domestic и MGCD International для 14 выбранных стран

Расстояния между национальными партнерами увеличиваются для Австралии, Японии и Южной Африки, тогда как расстояния с зарубежными партнерами снижаются. В случае Австралии и Южной Африки эта тенденция может отражать снижение зависимости от партнеров в Северном полушарии\*. Индия демонстрирует двойное снижение, где расстояния сотрудничества уменьшаются, как на национальном, так и на международном уровне. В глобализации научного мира эта модель отклонения может быть легко интерпретирована как структурная слабость индийской научной системы – неспособность индийских ученых идти в ногу с другими странами в осуществлении сотрудничества на длинных расстояниях. Однако более подробный анализ наших данных показывает, что этот результат, по крайней мере частично, получен благодаря изменениям в базе данных WoS. Оказывается, в период 2004 – 2010 гг. примерно 30 локальных индийских журналов, определенных нами просто как журналы со словом “Indian” («индийский») в их названиях, были внесены в базу данных. Скорее всего, поток публикаций из этих журналов сокращает расстояния между научными партнерами, поскольку эти публикации должны, главным образом, содержать адреса авторов в Индии.

## НЕОБХОДИМОСТЬ ДАЛЬНЕЙШЕГО ИССЛЕДОВАНИЯ

Научное сотрудничество, несомненно, является расширяющимся повсюду процессом в современной науке. Наши эмпирические результаты показывают продолжающийся рост географических расстояний между осуществляющими сотрудничество партнерами. По существу, это свидетельствует о том, что либо партнеры длинного расстояния участвуют в большей доле всех совместных связей, либо ученые устанавливают новые связи с партнерами, находящимися на более длинных расстояниях, чем было раньше. В анализе на макроуровне, приведенном в данной статье, мы не пытались разделить эти два процесса. Это требовало бы более подробного анализа на мезоуровне отдельных городов или на микроуровне отдельных научных учреждений.

Наш анализ также раскрыл заметную разницу между странами. Некоторые страны, несомненно, вносят вклад в научную глобализацию, а другие – нет. Последняя (вторая) группа может включать страны со значительным числом локальных журналов в базе данных WoS (журналы на местном языке или журналы на английском языке, предназначенные для локальных читателей). Наличие таких журналов в нашей базе данных могло повлиять на наши результаты, например, в случае Индии. Дальнейшие исследования должны учесть возможные влияния базы данных и, если необходимо, скорректировать их.

Кроме того, наши результаты обеспечивают достаточно свидетельство того, что процессы глобализации в науке продолжают и что действуют специфические для страны и области механизмы. Взаимодействие между этими двумя структурными факторами является интересной темой для будущего исследования. В данный момент мы можем только размышлять, наиболее вероятно, о характерных детерминантах наших основных ре-

зультатов. Нам бы хотелось предложить три темы для будущего эмпирического исследования:

- *Близость расстояния между научными партнерами.* Ученые, как и большинство людей, склонны к эффектам близости. Они предпочитают искать своих партнеров поблизости, с которыми их связывают один язык, культура и системы финансирования науки, что повышает вероятность успешного сотрудничества, результатом которого являются совместно написанные авторами публикации [11].

- *Стратегии и финансирование поддержки международного сотрудничества и участия в глобальных научных сетях.* Наличие таких структурных условий может поддержать или разрушить мотивацию среди специалистов, ученых, инженеров и техников, чтобы заниматься сотрудничеством на длинном расстоянии. Более подробные исследования, особенно на уровне стран или регионов, должны попытаться принять во внимание влияние национальных или наднациональных стратегических инициатив относительно нелокальных моделей и тенденций сотрудничества. Например, в случае Европы набирающая силу программа Европейской комиссии *Horizon 2020*, вероятно, будет крупным содействующим фактором в расширении внутриевропейского сотрудничества с распространяющимися эффектами относительно того, где европейские исследователи будут искать партнеров за пределами своего континента.

- *Концентрация ресурсов по науке и технологии в городах по всему миру.* Некоторые научные центры гораздо больше, чем другие [12, 13]. В нашей базе данных мы обнаружили, что менее 200 городов по всему миру, каждый из которых имеет крупные университеты и институты, достаточно для того, чтобы отвечать более чем за половину всей научной продукции. Благодаря десятилетиям или даже столетиям научного развития и накопления физического и человеческого капитала, такие центральные узлы в глобальной науке должны, по всей вероятности, привлечь соответственно большое число партнеров для сотрудничества.

## ЛИТЕРАТУРА

1. *Royal Society.* Knowledge, networks and nations: Global scientific collaboration in the 21<sup>st</sup> century. – London: The Royal Society, 2011.
2. *Cairncross F.* The death of distance: How the communications revolution will change our lives. – Boston: Harvard Business School Press, 1997.
3. *Hennemann S., Rybski D., Liefner I.* The myth of global science collaboration: Collaboration patterns in epistemic communities // *Journal of Informetrics.* – 2012. – Vol. 6, No. 2. – P. 217-225.
4. *Hoekman J., Frenken K., Tijssen R.J.W.* Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe // *Research Policy.* – 2010. – Vol. 39, No. 5. – P. 662-673.
5. *Jones B.F., Wuchty S., Uzzi B.* Multi-university research teams: Shifting impact, geography, and stratification in science // *Science.* – 2008. – Vol. 322. – P. 1259-1262.
6. *Wuchty S., Jones B.F., Uzzi B.* The increasing dominance of teams in production of knowledge // *Science.* – 2007. – Vol. 316. – P. 1036-1039.
7. *Tijssen R.J.W., Waltman L., Van Eck N.J.* Collaboration span 1, 553 kilometres // *Nature.* – 2011. – Vol. 473. – P. 154.

\* Южная Африка отличается от большинства других африканских стран, которые все еще очень полагаются на северное научное партнерство и связанный с этим приток ресурсов, чтобы поддержать или усовершенствовать свои еще не развитые научные системы [10].

8. *Waltman L., Tijssen R.J.W., Van Eck N.J.* Globalisation of science in kilometers // *Journal of Informetrics*. – 2011. – Vol. 5, No. 4. – P. 574-582.
9. *Liang L., Zhu L.* Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity // *Scientometrics*. – 2002. – Vol. 55, No. 2. – P. 287-316.
10. *Nordling L.* Africa analysis: Collaborating too far from home - Retrieved March 11, 2012. – <http://www.scidev.net/en/science-and-innovation-policy/south-south-cooperation/opinions/africa-analysis-collaborating-too-far-from-home-l.html>.
11. *Frenken K., Hardeman S., Hoekman J.* Spatial scientometrics: Towards a cumulative research program // *Journal of Informetrics*. – 2009. – Vol. 3, No. 3. – P. 222-232.
12. *Florida R.* The world is spiky // *The Atlantic Monthly*. – 2005, October. – P. 48-51.
13. *Leydesdorff L., Persson O.* Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes // *Journal of the American Society for Information Science and Technology*. – 2010. – Vol. 61, No. 8. – P.1622-1634.

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

Ответственный за выпуск      *Л. В. Кобзева*

Компьютерная верстка      *М. А. Филимонова*

ИД № 04689 от 28.04.2001 г.

Подписано в печать 28.06.2012 г.

Бумага «Хероx». Формат 60x841/8. Гарн. литер. Печать цифровая

Усл. печ. л. 5,00      Уч.-изд. л. 4,90      Тираж 55 экз.

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

Тел. (499) 155-44-95