



АНТОПОЛЬСКИЙ Александр Борисович -
доктор технических наук,
зав. лабораторией Института научной
и педагогической информации РАО (ИНИПИ РАО)
Адрес: 142432, Московская обл., Ногинский р-он,
пос. Черноголовка, Школьный б-р, 1
e-mail: ale5695@yandex.ru

КОЛЛАБОРАЦИОНИСТСКИЕ МЕТОДЫ ИЗМЕРЕНИЯ ИНФОРМАЦИОННЫХ РЕСУРСОВ

Введение

В настоящее время продолжается активный переход в электронную форму различных видов коммуникаций в научно-образовательной сфере. Здесь все большее значение приобретают электронные информационные ресурсы. Некоторые из них являются аналогами ресурсов, существовавших в доэлектронной (гутенберговской) информационной среде. Это книги, журналы, библиотеки, карты, учебные пособия, дидактические игры и проч. Другие появились в результате внедрения современных ИТ-технологий, прежде всего интернета. К ним относятся сайты, порталы, социальные сети, вики-ресурсы и многие другие.

В науке создание информационных ресурсов, в том числе в электронной форме, является основным измеряемым результатом. Появились и активно используются различные информационные инструменты, позволяющие получить количественные оценки этих результатов (публикационная активность, цитируемость, импакт-фактор, индекс Хирша и проч.). Обзор этих инструментов представлен, в частности, в работе автора [1] или в монографии [2]. Однако вопросы применения количественных, в том числе библиометрических показателей, для оценки эффективности научных исследований продолжают оставаться в центре внимания научной общественности. Много интересных публикаций на эту тему размещено, например, на сайте «Реорганизация Российской академии наук 2013» [3]. В данной статье эти вопросы непосредственно обсуждаться не будут.

Однако не вызывает сомнений то, что для многих задач, в том числе для сферы управления научно-образовательной деятельностью, имеется потребность в учете результатов информационной деятельности в научно-образовательной сфере, включающем количественные показатели, позволяющие объективно оценить эти результаты, то есть научно-образовательные ресурсы.

Одна из очевидных задач этого класса - учет электронных документных коллекций научно-образовательной сферы, которые могут быть объектом комплектования Национальной электронной библиотеки [4] или включены в состав Национальной системы хранения Интернет-ресурсов, создание которой предусмотрено Проектом основ государственной культурной политики [5].

Научно-образовательные ресурсы

Нужно заметить, что научно-образовательные ресурсы как объект исследований обладают заметной спецификой по сравнению с другими информационными ресурсами, как государственными, так и коммерческими.

Во-первых, в основном они публичные. Ресурсы ограниченного доступа (например, отражаю-

щие результаты научной деятельности в оборонной сфере или персональные данные ученых или учащихся) мы не рассматриваем.

Во-вторых, значительная доля этих ресурсов создана с использованием бюджетных средств, но не органами власти. Поэтому вопрос, являются ли они государственными, до сих пор не решен. Заметим, что автор является последовательным сторонником отнесения ресурсов, созданных с использованием бюджетных средств, к категории общественного достояния. Однако российский законодатель пока не разделяет эту точку зрения.

В-третьих, именно в научно-информационной сфере зародилось и активно развивается движение открытого доступа, которое является в значительной степени альтернативой коммерческому подходу к информационной деятельности. Подробнее об этом движении в [6].

Наконец, научно-образовательная сфера в силу специфики производителей и потребителей научно-образовательной информации является наиболее восприимчивой к технологическим новациям. Известно, что сам интернет (после того как был передан военными США для гражданского использования) быстрее всего распространялся в университетской среде. То же относится к другим ИТ, предназначенным для создания и использования информационных ресурсов.

Заметим в этой связи, что разделение научных и образовательных информационных ресурсов особого смысла не имеет, хотя, конечно, существуют собственно научные и собственно образовательные ресурсы. Однако пересечение этих категорий очень велико и, по экспертной оценке, составляет более половины всего научно-образовательного информационного пространства.

Таким образом, можно сделать вывод, что метрики для научно-образовательных ресурсов имеет смысл рассматривать как самостоятельные, хотя нужно иметь в виду опыт и методики измерения информационных ресурсов в других сферах. В то же время имеет смысл рассматривать научно-образовательные ресурсы как единое целое.

Статистические параметры информационных ресурсов

Анализ количественных показателей информационных ресурсов, видимо, стоит начать со статистических показателей, представленных в официальных российских и международных документах. Необходимо отметить, что возможности эти весьма ограничены. Так, в основополагающем документе «Стратегия развития информационного общества в Российской Федерации» [7] информационным ресурсам посвящено только 3 показателя:

- доля архивных фондов, включая фонды

аудио- и видеоархивов, переведенных в электронную форму;

- доля библиотечных фондов, переведенных в электронную форму, в общем объеме фондов общедоступных библиотек;

- доля электронных каталогов в общем объеме каталогов Музейного фонда Российской Федерации.

Международный статистический проект «Индекс экономики знаний» включает единственный показатель «Количество публикаций в научных и технических журналах на 1 млн жителей».

Государственная статистика в целом не отражает состояние информационных ресурсов страны. Так, в наиболее подробном статистическом исследовании развития информационного общества, осуществляемом Высшей школой экономики [8], информационные ресурсы фактически не отражаются. Единственный класс показателей, имеющих к ним отношение, - это набор экономических показателей организаций контент- и медиасектора, в том числе единственный натуральный показатель «Производство книг, газет и журналов». По разделу «Индикаторы науки» имеется набор показателей, отражающих публикационную и патентную активность [9]. Эти показатели, конечно, крайне недостаточно характеризуют научные информационные ресурсы. По разделу «Индикаторы образования» показатели, отражающие информационные ресурсы образования, отсутствуют полностью.

Все это вызывает большое недоумение. Ведь весь смысл понятия информационного общества заключается в новом качестве производства и потребления информации, прежде всего за счет новых видов коммуникаций. В то же время основной продукт и результат информационной деятельности - информационные ресурсы - фактически не является объектом государственного управления. Соответственно, отсутствует и статистика по этому результату, которая должна обеспечивать обратную связь в этом управлении.

Правда, в последние годы правительство начало предпринимать некоторые действия по активизации деятельности в области создания информационных ресурсов. Эти действия реализуются в форме «открытого правительства» и Совета по открытым данным при Правительстве РФ [10].

Единственная отрасль, где ресурсы находят хоть какое-то отражение в статистике, - это библиотечная отрасль. В недавно проводившейся всероссийской библиотечной переписи нашли отражение показатели **таблицы 1**.

Можно утверждать, что данные показатели совершенно недостаточно отражают состояние библиотечных информационных ресурсов. Поэтому автором в ходе исследования информатизации учрежде-

Таблица 1

ний культуры [11] был предложен более детальный набор показателей. Этот набор был экспериментально опробован в упомянутом исследовании, однако дальнейшего распространения не получил.

Нужно заметить, что в монографии автора «Информационные ресурсы России» [12] достаточно подробно рассматриваются статистические показатели, применяемые для оценки информационных ресурсов в целом и в различных сферах деятельности в частности, и делается вывод, что государственная статистика в этой области совершенно неудовлетворительна. Монография была написана более 10 лет назад, однако положение за это время не улучшилось.

В цитированной монографии предлагается следующее общее основание деления статистических показателей, связанных с информационными ресурсами: это показатели собственно ресурсов (связанные с формированием и хранением ресурсов) и показатели использования ресурсов (в библиотечном деле это называется показатели библиотечного обслуживания). Далее мы будем использовать это деление.

Следует отметить любопытный, хотя и весьма спорный перечень параметров информационных ресурсов, который предлагается в материале, к сожалению, анонимном, размещенном по адресу [13] на веб-сайте, посвященном веб-дизайну. Этот перечень следующий:

- содержание;
- охват;
- время;
- источник;
- качество;
- соответствие потребностям;
- способ фиксации;
- язык;
- стоимость.

Каждый из предлагаемых параметров подробно анализируется и для некоторых из них обсуждаются методы количественного измерения.

Однако данная попытка параметризации информационных ресурсов на сколько-нибудь общем уровне является скорее исключением.

Библиометрия

Пока основной дисциплиной, в рамках которой исследуются проблемы количественной оценки информационных ресурсов, является библиометрия. Здесь мы не будем подробно обсуждать биб-

1. Электронные ресурсы

	из них	
Объем собственных баз данных - всего тыс. записей	библиографических баз данных	в т. ч. объем электронного каталога

2. Формирование и использование библиотечного фонда

Наименование показателя	в т.ч. электронные издания
Состоит экземпляров всего на конец отчетного года	
Поступило экземпляров за отчетный год	
Выбыло экземпляров за отчетный год	
Выдано экземпляров за отчетный год	

3. Информационно-библиотечное обслуживание

Число посещений веб-сайтов данной библиотеки.

лиометрические подходы и методы, которым посвящено много полноценных публикаций. Из последних наиболее общих работ сошлемся на цитированную выше монографию С.В. Бредихина и А.Ю. Кузнецова, где в частности имеется прекрасная библиография.

Укажем только на важное ограничение: существующие методы библиометрии основаны почти исключительно на измерении научной периодики. Впрочем, в последние годы ведущие мировые библиометрические службы (Web of Knowledge, Scopus) стали учитывать и некоторые другие виды публикаций (монографии, труды конференций). Однако в целом научные и особенно образовательные ресурсы новых видов, очень широко представленные в интернете, эти службы учитывают совершенно недостаточно. Вообще библиометрические методы при всех их несомненных достоинствах подвергаются резкой критике. Основной мотив критики - это то, что применение формальных индикаторов для управления наукой согласно закону Гудхарта приводит к искажению смысла и качества научной деятельности. Закон Гудхарта заключается в том, что, когда социальный или экономический показатель становится целью для проведения социальной или экономической политики, он перестает быть достойным доверия показателем.

В последние годы наметилось движение отказа от использования библиометрических показателей, в частности от индекса цитируемости Web of Science, зафиксированное в принятом правительством Великобритании новом регламенте [14] оценки ученых и научных институций. При этом регламент допускает использование других инструментов, в частности Google Academia, по крайней мере, для некоторых научных отраслей.

Что касается России, то важным событием в области библиометрии в 2013 г. стало создание по заданию Минобрнауки РФ ресурса под названием «Карта российской науки» [15]. Поскольку этот ресурс задуман как объединение наиболее значимых российских и международных ресурсов, то его потенциал представляется весьма значительным. Однако существующее состояние этого ресурса таково (он находится в режиме опытной эксплуатации) и критика его столь убедительна [16], что реального использования «Карты науки» можно ожидать только после существенной модернизации этого ресурса. Одновременно возникает вопрос о перспективах РИНЦ после введения «Карты науки» в промышленную эксплуатацию и соотношении этих двух весьма солидных ресурсов.

Хотя библиометрия остается важным инструментом оценки научных результатов, ее ограничения становятся все очевиднее. Для компенсации этих ограничений, прежде всего связанных с развитием новых научно-образовательных коммуникаций, в последние годы стали разрабатываться новые сервисы, ориентированные на измерения ресурсов, представленных в интернете.

Проблема учета и классификации ресурсов

Самостоятельной проблемой, примыкающей к проблеме измерения ресурсов интернета, является проблема их учета. Действительно, для того чтобы измерять ресурсы, нужно определить, что такое научно-образовательные ресурсы, договориться об их классификации и определении каждого типа. Опыт показывает, что это очень непростая задача.

Существует довольно много различных каталогов научно-образовательных Интернет-ресурсов, однако они очень сильно различаются по охвату, полноте, точности и другим параметрам. В качестве примера приведем перечень каталогов электронных библиотек (ЭБ), которые являются одним из наиболее актуальных и полезных видов научно-образовательных ресурсов.

Одной из важных задач при выборе показателей для оценки является классификация ресурсов. Ясно, что для коллекций документов, например, электронных библиотек или периодических изданий должны использоваться иные показатели, нежели для отдельных документов, даже когда отдельный документ имеет размер, сопоставимый с размером коллекции. Точно также нельзя сравнивать текстовые и аудиодокументы, учебные, методические и научные материалы и т.п. Также может быть полезна дифференциация ресурсов по дисциплинам и уровню образования, например, так как это сделано на портале «Российское образование» на базе стандарта мета-данных RUSLOM.

Возможны и другие способы дифференциации, например, по новизне, по языку, используемым про-

граммным средствам, форматам и проч. В любом случае нужно стремиться, чтобы оценка производилась для однородных групп ресурсов. Однако последовательное применение глубокой классификации приведет к абсурду: когда все учтенные ресурсы окажутся несравнимы. Поэтому наряду с дифференциацией нужно иметь и механизмы обобщения ресурсов по различным основаниям.

Измерение ресурсов интернета

Далее мы рассмотрим сервисы, применяемые для количественной оценки Интернет-ресурсов. Таких подходов достаточно много.

Прежде всего, это сервисы, позволяющие оценить объемы Интернет-ресурсов как в физических единицах (например, в Гб), так и в логических, например, количество страниц или документов на сайте или количество файлов определенного типа. Существуют сервисы, подсчитывающие объем текстовых ресурсов, выраженные в словах, графических - в числе уникальных изображений, а также аудиовизуальных ресурсов (измеряется длительность).

Весьма многочисленны и наиболее подробно исследованы показатели, связанные с использованием информационных ресурсов в сети. Эти показатели обычно применяются для коммерческих целей: более востребованные ресурсы более привлекательны для рекламодателей.

Наконец, самостоятельными направлениями получившими распространение в последние годы, являются вебметрика, исследующая «видимость» сайта в интернете, и их связность на основе цитируемости веб-сайтов и системы гиперссылок, а также альтметрика, измеряющая упоминаемость информационных объектов в социальных сетях. Эти методы рассмотрим ниже.

Что касается измерения объема ресурсов российского интернета, наверное, приоритет в этой области имеет одна из самых известных российских Интернет-компаний - Яндекс.

Яндекс в течение ряда лет поддерживал сервис «Числа», в рамках которого подсчитывались следующие показатели.

Прямые величины:

- количество уникальных серверов (шт.);
- количество уникальных документов (шт.);
- объем проиндексированной информации, Гб.

Обратные величины:

- средний размер одной страницы (документа), Кб;
- среднее количество страниц на одном сервере, шт.;
- средний объем одного сервера, Мб.

Этот сервис позже был заменен несколькими новыми сервисами, которые обеспечивают сбор различной статистики по Рунету. В 2009 г. Ян-

декс выпустил интересный отчет по статистике контента Рунета [17], в котором содержались следующие данные:

- общее число сайтов Рунета;
- количество страниц на сайтах;
- распределение сайтов;
- объем текстовой, графической, аудио- и видеоинформации (заметим, что для разных видов информации Яндекс использует разные единицы измерения);
- показатели языка интернета (язык, статистика слов, орфографические ошибки, неологизмы и проч.).

В настоящее время Яндекс предлагает сервисы, позволяющие пользователю самому определить нужные показатели использования сайтов («Метрика») либо показатели использования слов в запросах («Wordstat») и получить необходимую статистику самостоятельно.

Статистика посещаемости сайтов

Как отмечено выше, в интернете активно развиваются сервисы, связанные с использованием информационных ресурсов, основанные на анализе статистики посещаемости сайтов и страниц. Эти сервисы называют также веб-аналитикой.

Веб-аналитика предназначена для следующих целей:

- развитие функциональности сайта на основании тенденций в поведении посетителей;
- оценка эффективности рекламных кампаний в интернете;
- выявление проблемных мест в структуре, навигации и контенте сайта.

На основе статистики посещаемости можно измерить:

- количество просмотренных веб-страниц;
- частотность ключевых слов, по которым посетители находят сайт в поисковых системах;
- географию посетителей;
- время, проведенное на веб-странице посетителем;
- переходы между веб-страницами;
- аудиторию сайта (случайные, постоянные посетители и т.д.).

Веб-аналитика - молодая отрасль не только в странах СНГ, но и в мире в целом. Однако организация Web Analytics Association уже ввела общепринятые стандарты, по которым могли бы производиться измерения и анализ веб-аналитики.

Википедия [18] следующим образом классифицирует инструменты веб-аналитики.

Анализаторы логов:

- WebTrends
- Webalizer
- AWStats.

Счетчики-рейтинги. Дают количество посетителей за день, неделю, месяц, за всю историю.

- Rambler's Top100
- Liveinternet
- Рейтинг@Mail.ru
- OpenStat
- HotLog.

Системы Интернет-статистики. Дают суммарную информацию по посещениям, выбранным по некоторому срезу, заданному пользователем.

- Google Analytics
- Piwik
- Яндекс.Метрика
- Liveinternet
- Рейтинг@Mail.ru
- OpenStat (бывш. Spylog)
- HotLog.

Системы Интернет-статистики с детализацией по просмотрам страниц. Кроме суммарной информации, дают информацию по просмотрам страниц внутри каждого посещения.

- Woopra (англ. Woopra).

Системы Интернет-аналитики с детализацией поведения посетителя на странице. Дают максимально возможную детализацию с возможностью просмотра всех действий посетителей: движений мыши, кликов, нажатий клавиш и т. д. По собранной поведенческой информации строятся отчеты в виде карт активности посетителей на странице.

- ClickTale
- Яндекс.Метрика
- SpyBOX.

Одними из наиболее популярных инструментов анализа и последующей оптимизации информационных ресурсов (как сайтов, так и отдельных его страниц) являются *индексы цитирования (ИЦ)* - некие обобщенные численные показатели ссылочной популярности (ссылочной цитируемости) страниц. ИЦ подсчитываются поисковыми системами для дальнейшего использования в алгоритмах ранжирования результатов. И если показатели ИЦ для сайтов в целом могут служить показателем их «раскрученности», то знание рангов отдельных страниц сайта позволяет определить как слабые, так и сильные стороны информационного ресурса. Далее приводятся несколько наиболее известных ИЦ.

Yandex mИЦ (CY) - тематический индекс цитирования поисковой системы Яндекс, особенностью которого является то, что он рассчитывается по отношению к тематически близким ресурсам.

Logarithmic Yandex mИЦ (LCY) - величина, вычисляемая из тематического индекса цитирования. Шкала LCY может изменяться от 0 до 100, что на данный момент соответствует тИЦ от 10 до 250000. У сайтов, лишенных или не имеющих тематического индекса цитирования, тИЦ равен -1. (© Сергей Холод, 2007).

Google PageRank (PR) - алгоритм расчета авторитетности страницы, используемый поисковой системой *Google*, основанный на методе вычисления веса страницы путем подсчета важности ссылок на нее. Шкала PR может изменяться от 0 до 10. Для большинства сайтов средней «раскрученности» PR равен от 4 до 5. Для очень хорошо «раскрученного» сайта PR равен 6. 7 - величина, практически недостижимая для множества сайтов, но иногда встречается. Значения 8, 9, 10 имеют исключительно популярные и значимые проекты. PR - это параметр, относящийся к каждой отдельной странице, но не ко всему сайту в целом. На одном сайте могут находиться страницы с разным PR. Для страниц, PageRank которым еще не назначен, рекомендуется возвращать величину - 1.

Alexa Traffic Rank (ATR) - место домена в списке, упорядоченном по частоте посещений домена, и *Alexa Rich Rank (ARR)* - место домена в списке, упорядоченном по доступности домена для посещений. Эти показатели рейтинговой системы *Alexa* (дочерней компании *Amazon*) основаны на подсчете частоты посещений и просмотра страниц сайта. Алгоритм подсчета трафика рейтинга *Alexa* очень прост и получается путем усреднения числа просмотров страниц в течение трех месяцев для данного сайта.

Комплекс-показатели

Site Rank (SR) - комплексный аналог ТИЦа на весь интернет, предложенный рекламными системами *XAP* и *TNX* и вышедший недавно на бета-тестирование. На SR влияют не только обратные ссылки, но и наличие проиндексированных страниц, ТИЦ, посещаемость, *Alexa Rank* и т.д. Разработчики надеются, что *Site Rank*, возможно, станет стандартом. Однако из-за такой многозначной комплексности недостаток становится большое время выдачи результата (от 5 до 120 секунд).

Приведем несколько примеров статистических сервисов.

Google Analytics [19] (сокращенно *GA*) - бесплатный сервис, предоставляемый *Google* для создания детальной статистики посетителей веб-сайтов. Статистика собирается на сервере *Google*, пользователь только размещает JS-код на страницах своего сайта.

Бесплатная версия ограничена 10 миллионами просмотров страниц в месяц. Пользователям с активным аккаунтом *Google AdWords* предоставляется возможность отслеживания неограниченного числа просмотров страниц. Особенностью сервиса является то, что веб-мастер может оптимизировать рекламные и маркетинговые кампании *Google AdWords* с привлечением анализа данных, полученных при помощи сервиса *Google Analytics*, о том, откуда приходят посетители, как долго они остаются на сайте и где они находятся географически.

Пользователи видят группы объявлений и отдачу

от ключевых слов в отчетах. Также доступны дополнительные возможности, включая разделение посетителей на группы. Пользователи сервиса могут определить цели и последовательности переходов. Целью может выступать страница завершения продаж, показ определенных страниц или загрузка файлов. Используя данный инструмент, маркетологи могут определять, какая из рекламных кампаний является успешной, и находить новые источники целевой аудитории.

Google Analytics показывает основную информацию «на панели инструментов», более детальную информацию можно получить в виде отчета. В данный момент доступно 80 видов настраиваемых отчетов.

«*Яндекс.Метрика*» [20] - бесплатный сервис, предназначенный для оценки посещаемости веб-сайтов и анализа поведения пользователей, открытый в 2009 г.

Счетчик «*Яндекс.Метрика*» работает по принципу обычного счетчика посещений: JS-код устанавливается веб-мастером на страницах сайта и собирает данные о каждом посещении.

Сервис интегрирован с *Яндекс.Директом* и *Яндекс.Маркетом* и позволяет группировать посетителя ресурса по нескольким параметрам.

Яндекс.Метрика измеряет конверсию сайта и Интернет-рекламы. При расчете конверсии сервис оценивает, какая доля посетителей сайта достигла «цели», то есть:

- дошла до некоторой страницы, посещение которой можно расценивать как достижение результата;
- просмотрела определенное количество страниц, которое является показателем успешности кампании;
- произвело определенное действие (клик на кнопку, скачивание прайс-листа и т. д.).

Счетчику «*Яндекс.Метрика*» может быть задано до 100 «целей». *Яндекс.Метрика* предоставляет данные за текущий день. Отчеты обновляются с периодичностью раз в 5 минут. Имеется функция мониторинга доступности сайта с возможностью получать SMS-уведомления, когда сайт недоступен пользователям.

Для создания детальных отчетов в сервис встроены специальный «Конструктор отчетов».

Для анализа последовательности просмотра пользователями страниц и разделов сайта *Метрика* предоставляет «Карту путей» по сайту (отчет показывает основные направления навигации пользователей по сайту и представляет собой граф, в котором вершины - это страницы или разделы сайта, а ребра - это пути пользователей).

«*Топ 100*» - сервис компании *Рамблер* [21] - предлагает для информационных ресурсов (сайтов), отображенных по заданному пользователем критериям, сводный показатель «индекс популярности», который на данный момент учитывает следующие аудиторские характеристики:

- количество хостов на главной странице за день;
- количество хостов на всем сайте за день;

- среднее суточное количество просмотров на хост за предыдущие 7 дней;
- количество хостов на главной странице за предыдущие 7 дней;
- количество хостов на всем сайте за предыдущие 7 дней;
- среднее число просмотров только 1 страницы за предыдущие 7 дней;
- среднее число просмотров от 3 до 5 страниц за предыдущие 7 дней;
- среднее число посетителей из России за предыдущие 7 дней.

Под хостом понимается уникальный IP-адрес. Индекс пересчитывается один раз в час. Кроме того, сервис предоставляет сведения о количестве уникальных посетителей, просмотра ресурса или его главной страницы за заданный период.

Google Academia

Отдельно следует упомянуть сервис Google Academia [22]. Во-первых, этот сервис рассчитан на научно-образовательные ресурсы, размещенные в интернете. При этом обрабатываются не только обычные периодические издания, размещенные в открытом доступе, но и определенные виды других категорий Интернет-публикаций, например, препринты, размещенные в открытых архивах. С другой стороны, он рассчитывает показатели, применявшиеся до того в традиционных библиометрических индексах, - публикационную активность и цитируемость, а также индекс Хирша. Таким образом, этот сервис пытается сочетать классические библиометрические показатели с новыми видами научных коммуникаций. Это, конечно, можно рассматривать как несомненное достоинство сервиса.

Нужно заметить, что использование Google Academia предлагается, например, новым правительственным регламентом по оценке научных учреждений Великобритании, цитированным выше. Правда, это допускается только для некоторых отраслей знания. Действительно, для многих дисциплин, в частности для педагогики, полнота отражения научно-образовательных ресурсов сервисом Google Academia значительно выше, чем в традиционных индексах цитируемости - Web of Science, Scopus или РИНЦ. Однако алгоритмы отбора ресурсов для индексирования этим сервисом неочевидны, поэтому результаты расчета показателей вызывают сомнения. Так или иначе, данный сервис следует рассматривать как один из наиболее адекватных для расчета, по крайней мере, некоторых показателей научно-образовательных ресурсов.

Вебометрика

О вебометрике и создании российского вебометрического индекса автор неоднократно писал, в том числе и в журнале «Информационные ресурсы

России» [23,24], поэтому подробно эта методология рассматриваться здесь не будет. Укажем лишь, что вебометрические показатели являются единственным инструментом, официально используемым для оценки научно-образовательных ресурсов интернета, конкретно, сайтов научных и образовательных учреждений. На этих показателях основан Вебометрический индекс университетов мира, эти показатели используются для формирования большинства известных университетских рейтингов. Напомним, однако, что классический вебометрический индекс строится на базе 4 показателей:

- объем сайта (число страниц);
- объем сайта (число так называемых «богатых» файлов);
- «видимость» сайта поисковиком Google Academia;
- цитируемость сайта (число ссылок на него).

Однако позже вебометрика стала использоваться и для других наукометрических задач. Например, А.А. Печников и его коллеги из Карельского научного центра РАН ведут вебометрические исследования связности научных и образовательных сайтов для кластеризации информационного научно-образовательного пространства [25] и построения общей модели российского университетского веба [26].

Альтметрика

Новым и пока малоизвестным в России инструментом оценки информационных ресурсов, к тому же специально ориентированным на научно-образовательную сферу, является сервис altmetrics, о котором мы расскажем более подробно, опираясь в основном на работу [27].

Выше мы отмечали, что использование библиометрических показателей, основанных на цитировании, особенно импакт-фактора, представляется многим исследователям неадекватным и слишком подверженным коммерческим и конъюнктурным факторам.

Все это, а также широкое распространение социальных медиа в научно-образовательной среде, привело к появлению альтернативной системы измерений научных информационных ресурсов, прежде всего статей и наборов данных, получившей название altmetrics (от альтернативной метрики).

Этот сервис, появившийся только в 2013 г., быстро завоевал популярность и сейчас используется многими научными учреждениями и издательствами. Сервис является агрегатором упоминания научного контента в СМИ, социальных сетях, блогах, постах и других Интернет-источниках. Существенно, что обрабатываются данные о конкретной статье, а не о журнале в целом, что делает оценку более точной. Впрочем, сервис позволяет, конечно, строить агрегированную оценку журнала, научного учреждения, отдельного учреждения или научного проекта.

Оценки при помощи альтметрики очень удобны для привлечения внимания к конкретным научным результатам, недостаточно отраженным в традиционных индексах цитирования, но привлечшим внимание в интернете.

Альтметрика представляет собой сводный показатель, который рассчитывается как сумма упоминаний в социальных медиа (Твиттер, Фейсбук и др.), а также упоминаний в специализированных научных блогах, сервисах Google+, Mendeleev и др.

Особенно альтметрика удобна для издателей, которые могут оценить внимание к публикациям в своих изданиях. Конечно, большое преимущество для применения альтметрики имеют публикации в изданиях открытого доступа, а также размещенные в открытых архивах. Сервис может быть также полезен при определении того, какие форматы онлайн-общения являются наиболее эффективными для распространения сообщений о научно-исследовательских достижениях.

По состоянию на ноябрь 2013 г. сервис собрал данные о 1,7 млн статей. Сервис, ориентированный на издателей, позволяет отбирать статьи по различным критериям, в том числе временным, по идентификаторам, ISSN, названию журнала и др. Эти данные собираются быстро и в больших количествах. Ежедневно сервис собирает 15 тыс. упоминаний о научных результатах; еженедельно служба фиксирует 22 000 уникальных статей (данные приводятся по состоянию на ноябрь 2013 года).

Дальнейшее развитие сервиса предполагает создание инструментов, обеспечивающих дифференциацию показателей, характеризующих ученых, научные учреждения в целом, результаты отдельных университетских лабораторий, исследовательских групп, а также межуниверситетских проектов.

Авторы подчеркивают, что показатели альтметрики существенно различаются для различных наук и дисциплин. Например, для биомедицинских наук они почти в четыре раза выше, чем для социальных. Предстоит уточнить, каким образом следует применять данный показатель для различных предметных областей.

Выводы и рекомендации

Анализ ситуации с измерением научно-образовательных ресурсов показывает, что показатели государственной статистики не отвечают современным потребностям ни в части создания ресурсов, ни в части их использования.

Методы библиометрии, применяемые в настоящее время для анализа и оценки информационных ресурсов в основном как результатов научной деятельности, пока занимают важное место в процессах управления наукой. Однако в перспективе, учитывая усиливающийся процесс перехода научно-образовательных коммуникаций в электронную фор-

му, следует ожидать повышения роли методов измерения и оценки научно-образовательных ресурсов, основанных на Интернет-технологиях.

В то же время мы видим, что для оценки и рейтингования информационных ресурсов, размещенных в интернете, применяются различные показатели: объемные, показатели посещаемости, видимость, цитируемость, альтметрика. Все они используют разнообразные инструменты, хотя для расчета большинства показателей, так или иначе, применяются самые распространенные поисковые сервисы - Google (мировые сервисы) и Яндекс (русские).

Центральной проблемой является поиск связи между количественными показателями информационных ресурсов и его качественными характеристиками. Особенно это существенно для научно-образовательных информационных ресурсов. Действительно, для рекламных целей оценка посещаемости ресурса, а также учет поведения пользователя на ресурсе могут достаточно точно аппроксимировать коммерческую ценность этого ресурса. Для научно-образовательных ресурсов это, конечно, не так. Их ценность определяется новизной, полнотой, качеством, что далеко не всегда коррелируется с посещаемостью. Например, наиболее посещаемыми в образовательной сфере являются банки данных рефератов, которые никак не могут признаваться наиболее качественными образовательными ресурсами.

Самостоятельной, но необходимой задачей при формировании базы данных характеристик научно-образовательных ресурсов является их учет. Учет может быть организован только на основе согласованной и общепринятой классификации этих ресурсов, включающей стандартизованные дефиниции выделенных типов.

Общий вывод можно сделать следующий: каждый из рассмотренных методов получения количественных показателей информационных ресурсов имеет свои преимущества и свою оптимальную сферу применения. Поэтому для общей системы мониторинга рассматриваемой сферы целесообразно сформировать сводную базу данных, где учтенные научно-образовательные ресурсы будут описаны набором метаданных, достаточно полно отражающих их специфику по признакам, релевантным для различных оценок ресурсов.

Далее для каждого ресурса следует иметь набор показателей, рассчитанных по различным методикам: объемные, веб-аналитические, вебометрические, альтметрические и др. База должна, естественно, обновляться автоматически с использованием адекватных сервисов, имеющихся в настоящее время в интернете. Понятно, что инструменты компании Google имеют значительный приоритет, но, наверное, могут быть не единственным используемым сервисом.

Возможно, в состав такой базы должны войти индикаторы, которые в настоящее время рассчиты-

ваются в составе Российского вебметрического индекса научно-образовательных учреждений. Как уже упоминалось, такой индекс в настоящее время ведется в Институте научной и педагогической информации (ИНИПИ) РАО.

Если предлагаемый метод будет в принципе одобрен научной общественностью и государственными органами управления наукой и образованием, можно обсуждать конкретный набор релевантных показателей.

Такой подход, который автор предлагает назвать коллаборационистским, позволит получать рейтинги научно-образовательных ресурсов по самым различным наборам признаков.

В качестве очевидных практических задач таких рейтингов можно указать на задачи отбора научно-образовательных ресурсов для комплектования Национальной электронной библиотеки, а также Национальной системы хранения Интернет-ресурсов.

Литература:

1. А. Антопольский. Использование информационных ресурсов для оценки эффективности научных исследований // Межотраслевая информационная служба. - 2011. - № 1. - С. 40-53.
2. С. Бредихин, А. Кузнецов. Методы библиометрии и рынок электронной научной периодики. - Новосибирск: ИВМиМГ СО РАН, НЭИКОН, 2012. - 256 с.
3. Реорганизация Российской академии наук 2013 [Электронный ресурс]. - URL: <http://www.saveras.ru>.
4. НЭБ: в приоритете расширение библиотечного охвата и экономическая эффективность [Электронный ресурс]. - URL: <http://mers.medart.tomsk.ru/news/194.html>.
5. Проект основ государственной культурной политики // Российская газета. - 16 мая 2014 [Электронный ресурс]. - URL: <http://www.rg.ru/2014/05/15/osnovi-dok.html>.
6. А.Б. Антопольский. Открытые данные в науке и образовании // Художественное образование в информационном обществе / Сб. научных трудов. - М.: МГУКИ, 2014. - С. 115-215.
7. Стратегия развития информационного общества в Российской Федерации от 7 февраля 2008. № Пр-212 // Российская газета. - 16 февраля 2008. [Электронный ресурс]. - URL: <http://www.rg.ru/2008/02/16/informacia-strategia-dok.html>.
8. Наука. Инновации. Информационное общество: 2012 / кр. стат. сб. - М.: НИУ «Высшая школа экономики», 2012. - 80 с.
9. Индикаторы науки: 2014 / статистический сборник. - Москва: НИУ «Высшая школа экономики», 2014. - 400 с.
10. Открытые данные 2014-2016 [Электронный ресурс]. - URL: <http://открытыеданные.большоеправительство.рф>.
11. А. Антопольский, С. Горушкина, О. Шлыкова. Состояние электронной культуры в отдельных регионах России: итоги комплексного исследования, проведенного в 2010 году / Справочник руководителя учреждения культуры, 2011. - № 5. - С. 6-21.
12. А. Антопольский. Информационные ресурсы России: научно-методическое пособие. - М.: Либерия, 2004. - 425 с.
13. Параметры информационных ресурсов [Электронный ресурс]. - URL: <http://elearn.oknetmian.ru/?p=6&id=117>.
14. Research Excellence Framework [Электронный ресурс]. - URL: <http://www.ref.ac.uk>.
15. Карта российской науки [Электронный ресурс]. - URL: <http://mapofscience.ru>.
16. Общество научных работников [Электронный ресурс]. - URL: <http://onr-russia.ru/content/карта-российской-науки>.
17. Информационный бюллетень Контент Рунета [Электронный ресурс]. - URL: <http://company.yandex.ru/researches/reports/yacontent09.xml>.
18. Основные показатели посещаемости сайта [Электронный ресурс]. - URL: <http://ru.wikipedia.org/wiki>.
19. Веб-аналитика корпоративного уровня [Электронный ресурс]. - URL: <http://www.google.com/analytics/>.
20. Веб-аналитика для повышения эффективности вашего сайта [Электронный ресурс]. - URL: <https://metrika.yandex.ru/list/>.
21. Rambler: Top100 [Электронный ресурс]. - URL: <http://top100.rambler.ru/>.
22. Google academia [Электронный ресурс]. - URL: <http://scholar.google.ru/scholar>.
23. А. Антопольский, Ю. Поляк, В. Усанов. О российском индексе веб-сайтов научно-образовательных учреждений // Информационные ресурсы России. - 2012. - № 4. - С. 2-7.
24. А. Антопольский, Ю. Поляк. О новом рейтинге веб-сайтов российских университетов // Дистанционное и виртуальное обучение. - 2013. - № 1. - С. 40-47.
25. А. Печников, Н. Луговая, Ю. Чуйко. О связности множества официальных сайтов РАН // Вопросы современной науки и практики. - 2009. - № 26. - С. 47-52.
26. А. Печников. Модель университетского Веба // Вестник Нижегородского университета им. Н.И. Лобачевского. - 2010. - № 6. - С. 208-214.
27. Jean Liu, Euan Adie. Realising the Potential of Altmetrics within Institutions // Ariadne Issue 72. - March 2014 [Электронный ресурс]. - URL: <http://www.ariadne.ac.uk/issue72/liu-adie>.