

ОСНОВЫ РАЗРАБОТКИ СИСТЕМЫ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ БИБЛИОТЕЧНОГО ВЕБСАЙТА

С.К. Канн, старший научный сотрудник
ГПНТБ СО РАН, Новосибирск

Постановка задачи. Современную ситуацию в библиотечном деле называют по-разному – «кризисной», «переломной», «революционной»... Суть заключается в том, что всё чаще библиотекарям приходится отвечать на трудные вопросы о будущем традиционных библиотек, о том, *«что с нами происходит»*, *что* и, главное, *как* делать. Как это ни трудно, приходится быстро меняться, меняя вместе с этим и окружающий мир.

Интернет ежедневно присваивает себе всё новые и новые функции, ещё недавно принадлежавшие исключительно библиотекам. Все инновации последнего времени, модные «библиобусы» и мобильные «КИБО» (комплексы информационно-библиотечного обслуживания¹), модернизация традиционно-предметной деятельности библиотек, не могут отменить наступления виртуальной *эры электронного чтения и электронного знания*. Создание библиотечных сайтов – лишь первый и, в значительной степени, стихийный шаг на пути в цифровое будущее – не только неведомое, но и непредсказуемое. Хочется надеяться, что стремительному продвижению вперёд будут приданы максимально сознательные формы, например, в сфере библиотечного сайтостроения.

Обзор большого количества библиотечных сайтов, а также публикаций в профессиональной печати убеждает в том, что библиотечной веб-статистике и аналитике, без которых сознательное сайтостроение попросту невозможно, уделяется недостаточное внимание. В лучшем случае, этим занимаются технические специалисты, обслуживающие «железо» и заинтересованные в определении критической нагрузки на сервер. Что касается информационных работников (контент-менеджеров, а также собственно библиотечных специалистов), то большинство из них всё ещё «страшно далеки» от названной темы. Относи-

¹ Интересно, что японское слово «Кибо» означает «Надежда». Такое имя было дано, например, японскому модулю Международной Космической станции.

тельные успехи достигнуты веб-аналитиками в области Интернет-маркетинга и рекламы, но их достижения обусловлены жёсткими реалиями экономического выживания и необходимостью оптимизировать каналы извлечения прибыли.

По словам Википедии, «веб-аналитика – молодая отрасль не только в странах СНГ, но и в мире в целом»². Тем не менее, уже существует ряд международных организаций, активно формирующих стандарты исследований и развивающих проблематику веб-анализа, например, DAA – Digital Analytics Association (бывшая *Web Analytics Association*). Осенью 2008 г. DAA предложила официальное определение понятия веб-аналитики как отрасли по «измерению, сбору, анализу и представлению Интернет-данных в целях понимания и оптимизации использования Интернета»³. Именно в таком виде формулировка вошла в англоязычное издание Википедии, но в отечественной Вики заменили слова «в целях понимания...» [*for purposes of understanding...*] на фразу «с целью улучшения», существенно исказив смысл нормативной формулировки DAA.



В действительности, существующие статистические средства и подходы к изучению веб-среды только на то и направлены, чтобы хоть как-то разобраться в сложнейших метаморфозах «всемирной паутины» (при этом, конечно, никто не против внедрения всяческих «улучшений»). Но без серьёзного понимания эволюционных процессов, происходящих в Сети, дальнейшее развитие знания невозможно. Следовательно, веб-аналитика, опирающаяся на широкое применение статистических методов, становится чрезвычайно важной как для современной информационной науки, так и для библиотековедения, как её части.

Отсутствие единообразно понимаемой статистики не сближает, а отдаляет библиотечные сайты друг от друга. На наш взгляд, потребности сопоставления

² Веб-аналитика // Википедия – свободная энциклопедия. URL: <http://ru.wikipedia.org/wiki/Веб-аналитика> (дата обращения: 20.09.2012).

³ Web Analytics Association Web Analytics Definitions – Draft for Public Comment 9/22/2008. Page 3.

URL: http://www.digitalanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitions.pdf (дата обращения: 20.09.2012).

данных, учёта текущего прогресса или регресса сайтов, регулярной отчётности учреждений в рамках управленческих взаимодействий, а также многие другие обстоятельства неумолимо предполагают:

- 1) унификацию используемой терминологии;
- 2) применение единообразных средств веб-статистики и
- 3) однозначную трактовку показателей (*e-метрик*).

Установление согласованной терминологии, общих критериев сбора веб-статистики и унифицированных инструментов веб-анализа позволит усилить эффективность управления библиотечными сайтами, укрепит взаимодействия в рамках единой информационной сети и создаст условия для совершенствования глобального разделения труда в сфере новых информационных технологий.

Обсуждение вопроса. Формирование системы показателей библиотечных сайтов, позволяющей отследить их внутреннюю динамику и сопоставить разные веб-метрики предлагается начать с установления более или менее *единообразной терминологии*. Решение этой задачи имеет особую актуальность в связи с повсеместным развитием электронных библиотек.

Несмотря на относительно протяжённый период существования библиотечного Интернета, до сих пор существует путаница понятий «сервер» и «сайт», «хит» и «страница», «доступ» и «посещение», «просмотр» и «визит». Немаловажно, что по своей технической структуре, по организации контента и прочим исторически сложившимся характеристикам сайты весьма индивидуальны, и это также не может не отражаться на терминологии, применяемой при описании их деятельности. Сблизить понимание предмета могут профессиональные дискуссии на библиотечных форумах и страницах печати.

Разный инструментарий, которым оснащены библиотечные сайты, разнообразные настройки программного обеспечения не только не гарантируют точности измерений, но не способны обеспечить простой сопоставимости результатов. Многолетний опыт показывает, что погрешности данных статистики могут достигать и 15, и даже 30%. Если бы появилась возможность оснастить биб-

лиотечные сайты унифицированными статистическими программами (модулями), смонтированными по единой методике на серверах библиотек, входящих в консорциумы, то это в серьёзной степени могло бы нивелировать статистические погрешности и предоставить всем желающим унифицированный экспериментальный материал для сопоставлений и выводов.

Сбор и интерпретация статистических данных осложняются рядом ограничений, связанных с техническими особенностями сети и конкретных сайтов, со множественностью подходов к сбору статистики, с наличием широкого спектра средств мониторинга и неопределённым («размытым») характером рассматри-



ваемых показателей. Казалось бы, для целей унификации и сопоставления данных можно было бы довериться системам «баннерной» статистики, образующим популярные рейтинги *Mail.ru*, *Rambler's Top100*, *OpenStat* (бывший *Spylog*), *HotLog*.

Утверждается, что в их основе лежат единые критерии и однообразно понимаемые «соревновательные» правила. Однако при ближайшем рассмотрении обнаруживается, что счётчики этих проектов нередко «измеряют» не реальную посещаемость ресурсов, а, вероятнее всего, технические условия доступа к сетям, особенности маршрутизации трафика и функционирования сайтов.

Наиболее дискуссионными, на наш взгляд, является ряд проблем.

1) Количественная сторона библиотечной веб-аналитики. Какие количественные показатели считать первостепенными, а какие менее существенными, в какой степени можно им доверять? Чем отличаются подходы к веб-аналитике в коммерческом секторе, веб-маркетинге и рекламе (показатели «отказов», «вовлечённости», «конверсии» и пр.), и некоммерческом (библиотечном)?

По нашему мнению, показатель «уникальных посетителей» принципиально важен именно потому, что *в значительно меньшей степени* подвер-

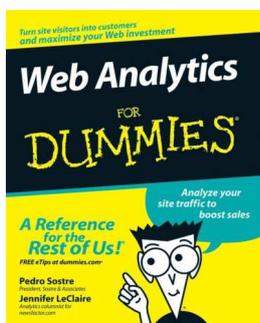
жен искажению со стороны автоматических средств индексации (роботов), чем цифра «обращений к страницам». Показатель «величины перекачки» тесно связан как с размерами индексации сайтов, так и с тяжестью размещённых на них файлов, поэтому принципиально важно публиковать *количественные данные о размерах сайтов*, об их файловой структуре и наполнении, *чего практически никто не делает*. Это также важно делать из-за прямой зависимости между количеством документов, представленных на сайте, и количеством обращений к ним, что требует вычисления особых «коэффициентов обращений» для сопоставления разнородных ресурсов.

2) Качественная сторона. Дискуссионным остаётся вопрос об учёте статистики роботов, так как очевидна прямая связь между индексированием сайта и популярностью («продвижением») ресурсов, веб-репутацией и посещаемостью страниц. Большой интерес для анализа содержательной стороны сайтов представляет и массив поисковых запросов, генерируемый поисковиками в результате предварительного индексирования документов. Не будем забывать также и о том, что роботы создают кэшированные копии страниц, дополняя возможности «прямого» обслуживания пользователей оригинальными страницами сайта.

Особый интерес вызывает отношение к иллюстративной информации и мультимедиа файлам (звучу, видео), так как в электронных библиотеках они являются существенным компонентом обслуживания и присутствуют в поисковых запросах и выдачах. Исключив из анализа статистику «картинок», мы рискуем потерять много значимой информации. Напомним, что многие электронные библиотеки предоставляют сканированные издания в форматах JPG или PNG. Из этого следует, что особенности учёта медиа-файлов представляют собой достаточно актуальную проблему.

Своего обсуждения и решения ожидают многие другие вопросы, как то: сведение воедино статистических показателей и создание комплексных отчётов по технологически разнородным информационным ресурсам (статическим документам и динамическим каталогам, базам данных); изучение особенностей

использования типологически разных электронных ресурсов (веб-навигаторов, библиографий, полнотекстовых материалов); критерии эффективности управления библиотечными сайтами; и многое другое.



Наши предложения. 1) Необходимо усилить внимание к развитию библиотечной веб-аналитики и вебометрики, причём не только со стороны *количественных* показателей создания и использования информационных ресурсов, структур и технологий, но и со стороны их *качественных* характеристик. Регулярный анализ качественных показателей использования

электронных ресурсов (веб-метрик) позволит более эффективно приспособлять работу Интернет-библиотекарей к быстро меняющимся условиям www-среды и возможной смене парадигмы информационной деятельности.

2) При оценке деятельности библиотечных сайтов следует учитывать оснащение электронных библиотек современными *инструментами веб-анализа* и представление подробных *статистических отчётов* для всеобщего ознакомления. Желательно, чтобы инструменты статистики носили унифицированный характер (как, например, система автоматизации библиотек ИРБИС) или, хотя бы, нормативно рекомендованный (из свободно распространяемого софта⁴).

3) Обозначилась потребность в создании *постоянной площадки* для обмена опытом в области библиотечной веб-статистики и аналитики, публикации материалов, обзоров и консультаций по названной проблематике. Трибуной для выступлений могли бы стать специализированные секции на регулярно проводимых конференциях РБА, ЛИБНЕТ, АРБИКОН и других библиотечных организаций, а также соответствующие рубрики или разделы в профессиональных изданиях («НТБ», «Библиосфера» и пр.).

⁴ По нашему мнению, на данный момент вполне приемлемой для этих целей является статистическая система AWStats, функционирующая на основе лицензии GNU General Public License. Последняя стабильная версия программы, которая монтируется на библиотечном сервере (ver. 7.0), появилась в декабре 2010 г., новая бета-версия 7.1 – 30.12.2011 г. Официальный сайт программы – <http://awstats.sourceforge.net>.