

УДК 027.5+004.738.5+004.414.22

С. К. Канн,

канд. ист. наук,

ст. науч. сотр. ГПНТБ СО РАН

ОЦЕНКА ДЕЯТЕЛЬНОСТИ БИБЛИОТЕЧНОГО САЙТА С ПОМОЩЬЮ СРЕДСТВ ВЕБ-АНАЛИТИКИ

В статье предлагаются подходы к оценке устойчивого развития и эффективной деятельности библиотечных веб-сайтов на основе использования современных средств и инструментов веб-анализа. Рассматриваются возможности применения маркетинговой концепции ключевых показателей эффективности учреждений (KPI) для оценки деятельности библиотечных сайтов. Освещаются некоторые проблемы библиотечной веб-статистики и аналитики.

Ключевые слова: библиотечный сайт, продвижение, посещаемость, веб-анализ, веб-статистика, система статистических показателей деятельности сайта, ключевые показатели эффективности.

Постоянный мониторинг текущих и долговременных показателей развития библиотечных сайтов необходим для улучшения их управляемости и повышения эффективности веб-обслуживания. Сама по себе задача анализа веб-трафика достаточно нетривиальна и сложна, что обусловлено многими обстоятельствами. Речь идёт об управлении очень большим массивом данных (*big data*), не всегда упорядоченных и зачастую малодостоверных. Свою лепту в нарастание энтропии вносит индивидуальность сайтов, специфика их контента, структуры и навигации. Из чего следует, что попытки свести процесс веб-анализа к самым простым решениям, скорее всего, будут безуспешными.

Тем не менее, накопленный опыт Отделения ГПНТБ СО РАН убеждает в возможности эффективного управления сайтом библиотеки, создания системы показателей его развития и взаимодействия с пользователями. Основу данной системы составляет деление всего объёма анализируемых данных о сайте на три блока показателей: 1) количественные данные о развитии контента; 2) показатели обращения к ресурсам (посещаемости) и 3) показатели веб-интеграции (веб-репутации, цитирования и пр.). Сопоставление основных цифр из разных блоков позволяет уверенно судить о прогрессе или регрессе сайта.

Никакие цифры посещений, сами по себе, не могут раскрыть специфики каждого отдельно взятого сайта, а тем более сопоставить его результаты с достижениями других сайтов. Опыт показывает, что количество внутренних и внешних обращений к ресурсам прямо пропорционально числу представленных на сайте документов (страниц). На посещаемости сайта отражается и физический размер его ресурсов, и общее количество файлов, и формат представленных документов. Значимыми характеристиками *контента*, подлежащими учёту, являются показатели соотношения текстовой и графической информации, средние цифры наполнения каждой html-страницы отдельными элементами («хитами»), количественные показатели хранящейся и предоставляемой мультимедиа-информации (звук, видео, флэш-ролики). Постепенно накапливаемый объём ресурсов создает прочный фундамент ежедневной посещаемости сайта, а соотношение количества хранящейся и отдаваемой информации характеризует эффективность его работы.

Индивидуальность каждого сайта связана с инструментами и методиками получения статистической информации. Собранные показатели (*e-метрики*) призваны отразить специфику организации ресурсов, наличие у сайта «зеркал» или кросс-доменных ресурсов, особенности контента (статический, динамический или смешанный). В зависимости от специфики каждого сайта используются различные инструменты и системы веб-аналитики.

К группе внутренних инструментов сбора веб-статистики относятся лог-анализаторы, обрабатывающие серверный журнал регистрации событий (лог-файл). Наиболее популярны бесплатные инструменты *analog*, *awstats*, *piwik*, *semonitor*, *webalizer*, *webtrends*. На сайтах, использующих преимущественно динамический контент и автоматические системы управления (CMS – content management systems), применяются специализированные модули: *1c-bitrix-web-analytics*, *drupal statistics*, *joomlastats*, *wordpress wp-statistics* и др. Обеспечивая значительную полноту данных об обращении к серверу, системы «внутренней» статистики, как правило, сильно завышают показатели посещаемости, так как не способны фильтровать обращения спам-ботов. Кроме того, нарастание размеров лог-файла периодически заставляет вести его усечение и архивацию.

На сайте Отделения ГПНТБ СО РАН www.prometeus.nsc.ru с 1998 г. используется «внутренняя» система лог-анализа – Advanced Web Statistics

7.3 (*awstats*), которая при всех своих достоинствах не умеет разделять данные по ботам, уникальным посетителям и доступу к ресурсам, а также по взаимному сопоставлению одного с другим. В процессе анализа нельзя изучить, какие обращения имеют интервалы доступа до 15 или до 30 секунд, и кому они принадлежат. Вследствие этого итоговая статистика «загрязняется» ботами, генерирующими, по приблизительной оценке, до 85 % трафика. Помимо этого выяснилось, что свободный доступ любого посетителя сайта к статистике *awstats* усиливает спамерские и хакерские атаки на сервер. Так, число доступов к *awstats* в 2014 г. по сравнению с предыдущим годом выросло почти в полтора раза и превысило показатели обращений к титульной странице сайта. После этого открытый доступ к статистике *awstats* был запаролен.

К внешним инструментам библиотечной веб-статистики и аналитики относятся счётчики-рейтинги (*hotlog*, *liveinternet*, *openstat*, *rambler's top100*), счётчики-трекеры (*яндекс.метрика*, *google analytics*) и различные системы интернет-аналитики, позволяющие детализировать информацию под разными углами зрения (*clicktale*, *iplogger*, *spybox*, *woopra*). Наиболее предпочтительным средством обработки данных о посещениях представляется *трекинг* сайта с использованием Яндекс-Метрики (далее: Я-М) и Гугл-Аналитики (GA). Обе системы хорошо фильтруют ботов по строгим правилам и по поведению, но и они не лишены существенных недостатков.

Механизм сбора и анализа статистики в названных системах основан на перехвате данных с помощью кодов *javascript*, внедрённых в страницы сайтов (так называемых страничных тегов). Служебная информация, «снятая» с помощью страничных тегов с браузеров пользователей библиотечных сайтов, передаётся на серверы Яндекса или Гугла, где и служит материалом для генерации отчётов. Любая ошибка в тегах или настройках скриптов на сайте библиотеки ведёт к искажениям и потере пользовательской статистики. Если брандмауэры посетителей, специальные браузеры (*epic privacy*, *pirate*, *tor*) или плагины (*disallow*, *ghostery*, *noscript*) препятствуют реализации *javascript*-кода, то это также нарушает целостность данных, как и сбои в любых сторонних скриптах. Наконец, около 3 % пользователей предпочитают вообще отключать реализацию скриптов. Нельзя не обратить внимание на постоянно растущее число продвинутых пользователей, применяющих приёмы, препятствующие отслеживанию (приватные режимы просмотра,

регулярную очистку файлов *cookie* и пр.). Таким образом, можно констатировать, что и внешние системы подсчёта веб-статистики далеко не безупречны.

При настоящем уровне технологий получить идеально точные данные абсолютно невозможно. Использование одних систем, обычно «внутренних», завышает показатели, других («внешних») – занижает. Таким образом, речь идёт о том, чтобы с учётом некоторых «доверительных интервалов» выяснить главные тенденции развития и обращения к ресурсам библиотечного сайта. Этому благоприятствует то обстоятельство, что корректно настроенные сайты отличаются большой устойчивостью показателей и, следовательно, величина ошибок остаётся примерно одинаковой на протяжении длительного времени. Это позволяет сравнивать многолетние данные за разные годы, вычислять их соотношение и динамику, делать прогноз на будущее.

В целом функционирование библиотечного сайта носит циклический и ярко выраженный сезонный характер. В «рабочий период» – с начала осени и до поздней весны – можно наблюдать всплеск посещаемости, причём, подобная тенденция наблюдается не только в библиотеках, но и на сайтах других культурно-образовательных учреждений, например кинотеатров [1]. Скорее всего, это связано с тождественностью веб-аудитории, охватывающей молодых людей в возрасте до 35 лет. Пик посещаемости библиотечных сайтов приходится на время студенческих сессий, а в летние месяцы происходит резкое падение показателей. В последнее время амплитуда сезонных колебаний значительно усилилась. Следует отметить и другую явно выраженную тенденцию, связывающую всплески посещаемости с регулярной актуализацией хорошо продвинутых ресурсов – выставок новых поступлений, библиографий, документальных и мемориальных коллекций. Подобную тенденцию отмечают авторы самых разных культурно-образовательных, научных и издательских проектов [1; 5, с. 176].

В количественных выводах библиотечной веб-статистики целесообразно опираться на концепцию, широко используемую в менеджменте и маркетинге, развитии бизнес-систем, торговли и пр. Главный подход заключается в том, чтобы учитывать *ключевые параметры эффективности* – так называемые *key performance indicators (KPI)*, не детализируя и не закликаясь на вторичных параметрах. Намечая конкретные *цели* по развитию KPI, можно постоянно двигаться вперёд,

наращивая эффективность создания и использования веб-ресурсов. При этом нельзя утверждать, что один набор данных – это «хорошо», а другой – «плохо». Просто показатели деятельности сайта надо воспринимать как некую «данность», то есть как *очень индивидуальные* метрики.

Например, опираясь на метрику «уникальных посетителей» за определённый период, более устойчивую к искажениям, чем «обращения» к сайту, можно расширять веб-аудиторию, вовлекая в круг пользователей новых посетителей и укрепляя «ядро» постоянных пользователей. Трекерные системы веб-аналитики, для которых наиболее существенное значение имеет параметр «сеансов» (или визитов), делят всех посетителей сайта на «новых» (*new visitor*) и «вернувшихся» (*returning visitor*). Если взять данные Гугл-Аналитикс за последний месяц, то у сайта Отделения ГПНТБ СО РАН (*prometeus.nsc.ru*) это соотношение составляет 72 % на 28 %; ГПНТБ СО РАН (*spsl.nsc.ru*) – 59 % на 41 %. При детальном рассмотрении ресурсов это соотношение может меняться, но в целом оно отражает *специфику* сайтов, а не их противопоставление по принципу «это – хуже, а это – лучше». Все параметры аналитики, среди которых показатели «отказов» (то есть доли визитов, ограниченных одной страницей), «глубины просмотра», «длительности сеансов», «конверсии» (доли целевых визитов) и пр., лишь описывают индивидуальную ситуацию каждого сайта, позволяя улучшать определённые *ключевые параметры* (KPI).

Третий блок в системе статистических показателей библиотечного сайта касается интеграции в веб-среду. Для оценки успешности продвижения сайтов в Интернет-пространстве используются специализированные инструменты веб-аналитики, аккумулирующие данные о количестве прямых и обратных ссылок, позиции сайтов в каталогах, рейтингах и поисковых системах, упоминаниях в социальных сетях. Как известно, количество ссылок с других сайтов, а также статистика цитирования полнотекстовых материалов положены в основу различных проектов испанской исследовательской группы Cybermetrics Lab по вебметрическому ранжированию крупнейших университетов и научно-исследовательских учреждений мира [8].

Большое количество е-метрик собирают специализированные сайты SEO-анализа и поисковой оптимизации, такие как: *pr-cy.ru*, *xtool.ru*, *linkpad.ru*, *majesticseo.com* и др. В отношении библиотек полученная информация может быть использована для оценки веб-репутации и

авторитетности сайтов, выявления наиболее востребованных ресурсов и перспективных тематик (ключевых слов, поисковых фраз и пр.). Как правило, системы SEO-анализа используют специально разработанные интегрированные индексы, такие как взвешенный индекс цитирования (ВИЦ), индекс Х-траст от компании Xtool, тематический индекс цитирования Яндекса (ТИЦ), PageRank (PR), применяемый Гуглом для обозначения уровня авторитетности ресурсов, и др.

С помощью популярной платформы *SimilarWeb*, кнопку которой можно смонтировать прямо в браузере, появляется возможность анализировать и сравнивать ключевые показатели посещаемости, ежемесячного объёма трафика, глубины и длительности просмотра страниц по каждому сайту. *SimilarWeb* позволяет структурировать источники посещений в разрезе прямого, ссылочного и поискового трафика, подсчитывать заходы из социальных сетей, почтовых рассылок и баннерных сетей. С помощью платформы можно выявить весь круг ссылок, связывающих библиотечный сайт с его партнёрами и аналогичными сайтами. Пример группировки данных приводится в табл. 1.

Таблица 1

Сравнение индикаторов развития библиотечных сайтов
с помощью аналитической платформы *SimilarWeb*
(данные на 30.06.2016 г.)

Библиотека	Веб-адрес	Посещаемость за месяц				Источники трафика за последние три месяца		
		Визитов тыс.	Время на сайте	Глубина просмотра (страниц)	Доля отказов	Поиск	Ссылки	Соцсети
Российская государственная библиотека	rsl.ru	686	4:33	6,60	37,7%	48,0%	22,0%	2,1%
Национальная библиотека Украины	nbuv.gov.ua	485	1:42	2,01	55,9%	80,1%	5,4%	1,4%
Российская национальная библиотека	nlr.ru	400	4:40	7,03	42,9%	49,2%	22,0%	3,1%
ГПНТБ СО РАН	spsl.nsc.ru	189	1:17	2,13	53,2%	83,3%	5,3%	1,0%
Национальная библиотека Беларуси	www.nlb.by	113	2:28	2,95	48,9%	80,2%	8,4%	0,7%
Отделение ГПНТБ СО РАН	prometeus.nsc.ru	56	2:19	2,23	58,3%	74,3%	6,0%	0,7%
Национальная библиотека Республики Казахстан	www.nlrk.kz	25	3:56	7,20	42,0%	84,0%	8,4%	0,4%

С этой точки зрения важно подчеркнуть нарастающее внимание систем веб-аналитики к взаимодействию сайтов с социальными сетями, потому что за этим взаимопроникновением можно увидеть ближайшее будущее «очередного этапа развития библиотеки как социального института» [6, с. 107]. Переходы на библиотечный сайт из социальных сетей, как правило, свидетельствуют об интересе к библиотечным ресурсам со стороны наиболее продвинутой молодой аудитории пользователей, опирающейся на самые современные средства мобильного интернета (смартфоны, планшеты и пр.). Трекерные системы веб-аналитики позволяют количественно оценить этот источник трафика.

На сайте Отделения ГПНТБ СО РАН тенденция к росту мобильных источников трафика носит очевидный характер. Доля мобильных посещений с 2011 по июнь 2016 г. возросла с 1,8 % до 11,1 % (табл. 2).

Таблица 2

Доля визитов к сайту *www.prometeus.nsc.ru*,
приходящихся на мобильные и немобильные устройства
(по данным Яндекс-Метрики за 2011–2016 гг.)

	2011 (с 14 марта	2012	2013	2014	2015	2016 (на 30 июня
всего визитов	604 385	587 481	434 585	375 457	391 614	134 327
Настольные компьютеры	98,2 %	97,1 %	95,8 %	92,2 %	89,1 %	88,9 %
Мобильные устройства и КПК	1,8 %	2,9 %	4,2 %	7,8 %	10,9 %	11,1 %

Поскольку по самым разным подсчётам главными пользователями сайта *prometeus.nsc.ru* являются молодые люди в возрасте до 24 лет (в том числе: 37,6 % среди мужчин и 50,1 % среди женщин), в дальнейшем можно ожидать увеличения мобильных посещений. По исследованиям международной аналитической компании GfK проникновение Интернета среди молодежи (от 16 до 29 лет) достигло почти предельного значения – 96 %, что в ближайшей перспективе означает назревшую необходимость для всех библиотечно-информационных структур ориентироваться на эту быстро растущую группу.

Подводя итоги, отметим, что современные проблемы библиотечного дела, и библиотечного Интернета в частности, связаны, прежде всего, с высокой изменчивостью веб-среды, в которой ежесекундный прогресс технологий стал обычным явлением. Глубокое противоречие заключается в том, что библиотеки всегда стремились к устойчивому развитию, опирались на прочные, исторически сложившиеся реалии, годами наработанные схемы и классификации. В настоящее время они переживают один из самых важных моментов своей многовековой истории, приобщаясь к непрерывной и всё более ускоряющейся информационной революции.

В этих условиях библиотечная веб-аналитика призвана сделать переход библиотек к новому состоянию более плавным и предсказуемым. Регулярный сбор, измерение и анализ вебметрических показателей позволяют рационально проектировать и развивать ресурсы, оптимизировать внутреннюю структуру и навигацию библиотечных сайтов, откликаясь на запросы пользователей. Можно определить направления будущего развития и лучше подготовиться к грядущим изменениям. Таким образом, реальные, а не мнимые успехи библиотечной веб-аналитики позволят максимально полно реализовать предназначение библиотек в современном информационном обществе.

Список использованных источников

1. Александрова У. А. Анализ статистики посещаемости сайта типичного красноярского кинотеатра / У. А. Александрова, С. И. Сенашов // Актуальные проблемы авиации и космонавтики. – 2014. – Т. 1, № 10. – С. 356.
2. Вирин Ф. Ю. Пути аудитории по сайту / Ф. Ю. Вирин // Интернет-маркетинг. – 2004. – № 1. – С. 30–39.
3. Гендина Н. И. Официальный веб-сайт: проблемы отражения в открытом информационном пространстве основных функций библиотеки как социального института / Н. И. Гендина, Н. И. Колкова, О. И. Алдохина // Науч. и техн. б-ки. – 2010. – № 2. – С. 5–15.
4. Гуськов А. Е. Альтернативная вебметрика: исследование веб-трафика сайтов научных организаций / А. Е. Гуськов, Е. С. Быховцев, Д. В. Косяков // Науч.-техн. информ. Сер. 1: Орг. и методика информ. работы. – 2015. – № 12. – С. 12–28.
5. Денисов А. А. Анализ статистических данных в управлении веб-сайтом на базе WordPress.com (на примере сайта журнала «Перспективы науки и образования») / А. А. Денисов // Перспективы науки и образования. – 2014. – № 5 (11). – С. 173–184.

6. *Жабко Е. Д.* Среда электронных библиотек: новые возможности повышения социальной активности пользователей / Е. Д. Жабко // Вестн. Санкт-Петерб. гос. ун-та культуры и искусств. – 2010. – № 1. – С. 105–108.
7. *Зобнин Б. Б.* Модель статистики посещаемости сайта / Б. Б. Зобнин, А. Н. Овчинников // Автоматизированные технологии и производства. – 2013. – № 5. – С. 114–116.
8. *Исмагилова А. Х.* Оценка качества электронных библиотек / А. Х. Исмагилова // Науч. и техн. б-ки. – 2010. – № 5. – С. 60–66.
9. *Кабакова Е. А.* Веб-сайт научно-исследовательского учреждения: наполнение, посетители, развитие / Е. А. Кабакова, В. С. Усков // Вопросы территориального развития. – 2014. – № 3 (13). – С. 4.
10. *Мухамедьянова Д. Н.* Анализ механизмов сбора статистики и продвижения сайта в сети Интернет / Д. Н. Мухамедьянова // Теоретические и прикладные вопросы науки и образования : сб. науч. тр. по материалам междунар. науч.-практ. конф., Тамбов, 30 сент. 2015 г. : в 16 ч. – Тамбов, 2015. – Ч. 11. – С. 79–80.
11. *Печников А. А.* О вебметрическом индикаторе «размер сайта» / А. А. Печников // Обзорение прикладной и промышленной математики. – 2013. – Т. 20, № 4. – С. 568.
12. *Печников А. А.* Об измерениях вебметрических индикаторов / А. А. Печников // Международный журнал экспериментального образования. – 2013. – № 10–2. – С. 400–404.
13. *Печников А. А.* Размышления о вебметрическом рейтинге / А. А. Печников // Научная периодика: проблемы и решения. – 2014. – № 1 (19). – С. 17–21.
14. *Поляк Ю. Е.* К измерению вебметрических индикаторов / Ю. Е. Поляк // Компьютерные измерительные технологии : материалы I Междунар. симпозиума, 3 апр. 2015 г., Москва. – Москва, 2015. – С. 293–296.
15. *Поляк Ю. Е.* Российский и международный опыт вебметрических исследований / Ю. Е. Поляк // Информационные ресурсы России. – 2014. – № 6 (142). – С. 2–9.
16. *Пронина Л. А.* К проблеме продвижения веб-сайтов общедоступных библиотек / Л. А. Пронина, Н. Д. Романова // Гаудеамус : психол.-пед. журн. – 2013. – № 2 (22). – С. 77–81.
17. *Скородумов П. В.* Анализ подходов и инструментальных средств анализа статистики посещения веб-сайта научной организации / П. В. Скородумов, А. Ю. Холодев // Вопросы территориального развития. – 2015. – № 9 (29). – С. 6.
18. *Торлин И. Г.* Ресурсы веб-сайтов публичных библиотек Украины: компромисс предлагаемого и ожидаемого / И. Г. Торлин // Науч. и техн. б-ки. – 2008. – № 3. – С. 36–41.
19. *Хайбуллин Р. Р.* Продвижение web-сайтов библиотек в среде Интернет / Р. Р. Хайбуллин // Вестн. Казан. гос. ун-та культуры и искусств. – 2007. – Спец. номер. – С. 301–303.

Статья поступила 01.08.2016

Sergei Kann

EVALUATION OF LIBRARY SITE' ACTIVITY USING WEB-ANALYTIC MEANS

The paper offers an approach to the evaluation of sustainable development and efficient operation of library websites using modern web analytical tools. The possibilities of application of the marketing concept of key performance indicators of institutions (KPI) for the evaluation of library sites activities and some problems of library web statistics and analytics are discussed as well.

Keywords: library site, promotion, attendance, web-analysis, web-statistics, statistical indexes system, key performance indexes (KPI).