

ПОСЕЩАЕМОСТЬ РЕСУРСОВ БИБЛИОТЕЧНОГО САЙТА ПО ОЦЕНКЕ GOOGLE ANALYTICS

Канн С.К.

Государственная публичная научно-техническая библиотека СО РАН, Новосибирск

В сообщении изложены результаты исследования ключевых показателей эффективности библиотечного сайта Отделения ГПНТБ СО РАН. С помощью системы Google Analytics собраны и проанализированы данные об аудитории и посещаемости библиотечно-библиографических ресурсов за 18 месяцев. Выявлены специфические особенности веб-метрик библиотечного сайта с учётом его информационной природы, политематичности представленных материалов и типологических особенностей ресурсов. Разработаны рекомендации для продвижения библиотечных сайтов в веб-среде.

Ключевые слова: Google Analytics, библиотечный сайт, система статистических показателей деятельности, посещаемость, веб-аналитика, ключевые показатели эффективности.

LIBRARY WEBSITE TRAFFIC ON THE ESTIMATION OF GOOGLE ANALYTICS

Kann S.K.

State Public Scientific and Technical Library SB RAS, Novosibirsk, Russian Federation

The report outlines the results of KPI study of the library website of the Branch of the SPSTL SB RAS. Using the Google Analytics, data on the audience and attendance of library and bibliographic resources for 18 months have been collected and analyzed. Specific features of the library website metrics have been revealed, taking into account its information nature, the polythematism of the materials presented and the typological features of the resources. Recommendations for promoting library sites in the web environment are developed.

Keywords: Google Analytics, library website, system of statistical indices, attendance, web-analytics, key performance indices (KPI).

Введение. Современные системы веб-трекинга и аналитики, такие как Яндекс-Метрика (ЯМ) или Google Analytics (GA), позволяют существенно улучшить эффективность управления сайтами. Проблема, однако, состоит в том, что стратегия продвижения ресурсов в значительной степени зависит от индивидуальности каждого сайта. Библиотечные сайты, как нам кажется, имеют свою ярко выраженную специфику. В данном исследовании ставилась задача проанализировать трафик библиотечного сайта Отделения ГПНТБ СО РАН *prometeus.nsc.ru* с точки зрения выявления его своеобразия, особенностей посещаемости, вытекающих из информационной природы, политематичности представленных материалов и типологических особенностей ресурсов – библиографических указателей, справочников, веб-навигаторов, полнотекстовых документов. В качестве главного инструмента анализа использовалась система Google Analytics, а полученные результаты сопоставлялись с данными, извлечёнными из трекинга Яндекс-Метрики и системы лог-анализа AWStats (Advanced Web Statistics 7.3), смонтированной

на сервере Отделения.

Статистические показатели разных сайтов могут отличаться в очень значительной степени. Так, например, показатели *лендинга* («посадочной страницы») имеют свою специфику, весьма отличную от показателей блога, а индикаторы *информационного* (библиотечного, некоммерческого) сайта далеко не тождественны показателям интернет-магазина. Наиболее наглядно эту разницу можно показать на примере метрики *отказов* (см. *Таблицу 1*).

Таблица 1. Средний уровень отказов на сайтах, имеющих разный тип контента [1].

Тип контента	Показатель отказов, %			
	Плохой	Нормальный	Хороший	Замечательный
Блог	80	70–80	60–70	Менее 60
Информационный сайт	Более 80	70–80	60–70	Менее 60
Интернет-магазин, крупный интернет-портал с различным контентом	Более 40	35–40	20–35	Менее 20
Коммерческий сайт с каталогом продукции	Более 55	45–55	35–45	Менее 35
Лендинг, посадочная страница	До 100	Менее 90	Менее 80	Менее 70

Использование целого ряда аналитических инструментов позволяет получать более достоверные и надёжные результаты в условиях большого разнообразия и сложности веб-среды. По замечанию известного специалиста Авинаша Кошика «каждый из применяемых в нашем мире инструментов страдает своим уникальным набором недостатков – от способа реализации кода до таких особенностей, что инструмент не сможет выполнять сегментирование при отсутствии предварительно определенных параметров и устанавливаемых пользователем переменных» [2, с. 463]. К этой мысли присоединяются и авторы «Web Analytics Strategies for Information Professionals», которые указывают, что «для библиотек, имеющих одинаковые кадровые и временные ограничения, установка нескольких платформ веб-аналитики является стратегией, нацеленной на сбор значительных объёмов данных без проведения сложной настройки и конфигурирования инструментов» [5, р. 149].

Недостатки инструментария. Внешние сетевые решения, подобные Google Analytics и Яндекс-Метрике, относятся к типу SaaS, т. е. software as a service («программное обеспечение как услуга»). Они основаны на внедрении JavaScript-кода («страничных тегов» и веб-маяков) в страницы сайтов и последующем перехвате служебной информации веб-браузеров. Точность сгенерированных отчётов зависит от массы автономных факторов: 1) ошибок в тегах и настройке скриптов, ведущих к необратимой потере данных; 2) ограничений брандмауэров на стороне клиента; 3) невозможности отследить окончание загрузки страниц, когда код реализуется в момент запроса; 4) неточной идентификации ботов и спамеров; 5) растущей анонимизации пользователей, блокирующих скрипты и очищающих файлы *cookie*. Подмена IP-адресов на прокси-серверах в значительной степени искажает не только географические, но и количественные индикаторы веб-аудитории.

В целом, погрешности счёта Google Analytics оцениваются нами в пределах 5-10% в сторону занижения. Есть свидетельства, что метрики, полученные с помощью GA, занижаются ещё больше – на 25% [4, с. 8, 10]. На наш взгляд, это может происходить из-за жёстких ограничений и собственных подходов GA к регистрации отказов. Вместе с тем, качественная фильтрация бот-трафика позволяет использовать Google Analytics как наиболее адекватное средство анализа. Весомым преимуществом GA также является удобный, интуитивно понятный интерфейс, совмещённый с высокой функциональностью. Google Analytics позволяет рассматривать данные под разными углами зрения, сегментировать аудиторию и даже мониторить посещаемость сайта в режиме реального времени. В силу названных причин Google Analytics использовался в качестве основного инструмента исследования, а другие системы (Яндекс-Метрика, LiveInternet, AWStats) – в качестве дополнительных.

Подчеркнём, что в последние годы уровень *реферального спама* резко возрос. Это напрямую связано с развитием автоматизированных технологий и усилением их влияния на сеть. По наблюдению А.А. Леонтьева, с которым нельзя не согласиться, «инструмент анализа общей статистики посещений даёт, прежде всего, картину столкновений и борьбы неких автоматических технологий на фронте компьютерных, т. е. тоже автоматических, информационных технологий» [3, с. 34]. Без сомнения, прогресс технологий может вызывать не только положительные, но и разрушительные эффекты. Придётся признать, что, занимаясь наращиванием библиотечных ресурсов, мы проглядели момент резкого усиления активности спам-ботов и «чёрных SEO-оптимизаторов», хотя первые признаки негативных явлений на сайте Отделения ГПНТБ СО РАН возникли, как теперь понятно, уже в 2005 году – в связи с усилением трафика к отдельным «популярным» файлам и отчётам AWStats.

Популярные, давние и устойчивые библиотечные ресурсы стали подвергаться автоматическим атакам ботов с весьма примитивной целью накрутить ссылочный вес рефереров. Постепенно ресурсы становились объектом манипуляций SEO-оптимизаторов, а в логи сервера записывалась информация, загрязнявшая отчёты. Система AWStats, обрабатывая журнал, не способна отсеивать «мусор». Мало того, будучи открытой для внешних посещений, она сама подвергается многочисленным атакам. При этом спам-адреса фиксируются во вкладке «Рефереры → Ссылающиеся сайты». URL записанных рефереров не имеет ничего общего с библиотекой и, в лучшем случае, восходит к сервисам бытовой техники, интернет-магазинам модной одежды и порталам диетического питания, а в худшем – к сайтам далеко не культурного содержания. К сожалению, нам неизвестны способы автоматической блокировки спам-ботов. Когда данная проблема обозначилась достаточно чётко, мы вручную стали вносить ip-адреса самых наглых ботов в чёрный список файла *htaccess* («*deny from...*»). На 30.09.2017 в список «забаненных» попало уже 5625 адресов. Помимо прочего регулярное отслеживание и удаление спама приносит положительный эффект в том отношении, что улучшает репутацию сайта и поднимает его позиции в поисковых выдачах Яндекса и Гугла.

AWStats помогает отслеживать бот-трафик, так как один из отчётов «Хосты → Полный список» позволяет сравнить метрики *обращений* и *«хитов»*, т. е. отдельных

(составных) элементов страниц. Значительная близость или полная тождественность этих двух показателей напрямую указывает на ботов, позволяя оценить их немалый вклад в структуру трафика. В 2015–2016 гг. было установлено, что на библиотечном сайте *prometeus.nsc.ru* доля ботов достигает 84–88% всех доступов. Активное использование «бана», обрезающего спам, понизило долю ботов до 75–80%. Но остальную часть трафика – в пределах 20–25% – можно считать «эффективной», т. е. направленной на обслуживание реальных пользователей.

Если проанализировать сводку основных показателей посещаемости (см. Таблицу 2), можно заметить, что показатели просмотров по отчётам Google Analytics и Яндекс-Метрики составляют 26,5% и 29% от уровня обращений, зафиксированных системой AWStats. Этот факт косвенно подтверждает правильность подхода к модерации, учитывая, что трекинг-системы всегда и с разной степенью успешности избавлялись от ботов. Вообще, внешние системы аналитики (GA, ЯМ, LiveInternet и др.) обычно занижают статистику «живых» пользователей, тогда как системы внутреннего, серверного лог-анализа всегда её сильно завышали. По названной причине системы подобные AWStats могут быть рекомендованы только для оперативного управления сайтом.

Таблица 2. Сводка основных показателей посещаемости сайта *www.prometeus.nsc.ru*, полученных разными инструментами аналитики за период 01.03.2016 – 31.08.2017 *

	Визиты	Посетители	Просмотры	Новые посетители	Отказы	Глубина просмотра	Время на сайте
Google Analytics	374417	273409	691744	273323	78,5 %	1,85	1:41
Яндекс-Метрика	406491	293312	759465	292465	77,9 %	1,87	1:54
AWStats 7.3	874298	556688	2611226	–	–	–	–

* Прочерк означает отсутствие данных в системе.

Анализ трафика. Для выяснения специфических особенностей трафика был взят период стабильных показателей, отражающих деятельность библиотечного сайта за 18 месяцев (01.03.2016 – 31.08.2017 г.). Сюда вошли как данные «рабочего периода» с высокой посещаемостью (осень-весна), так и статистика летнего затишья. На основе этих цифр Google Analytics сформировал ряд отчётов, касающихся контента, аудитории пользователей и др.

Отметим характерную особенность, бросающуюся в глаза при первом же знакомстве с обзором GA – высокую долю новых посетителей, до 97–99% по просмотрам страниц. Метрика «новых сеансов» (*new visitors*) за анализируемый период составляет 72,94%, а повторных посещений (*returning visitors*) – лишь 27,06%. В то же время уровень отказов в группе новых посетителей составляет 82,75%, а у возвращающихся 66,85%, при среднем показателе для представления 78,46%. Так как основным источником визитов на сайт Отделения является органический поиск (78,4% по данным GA), то, естественно, подавляющая часть пользователей, попадая на документы прямо из поисковой выдачи, тут же их и покидает. Иное дело – «возвратные» посетители, часть из которых со временем формирует устойчивое ядро постоянных пользователей. Глу-

бина просмотра у этой части аудитории превышает метрику новичков ровно в два раза – 2,92 страницы за сеанс против 1,45 (при средней 1,85).

С одной стороны, «новизна» сеансов GA тесно связана с размещением файлов cookies на компьютерах пользователей. Если cookies по каким-либо причинам не устанавливаются или регулярно очищаются, то и доля новых пользователей, очевидно, стремится к 100%. Вместе с тем, представленные показатели больше всего характеризуют устойчивость и неизменность библиотечного контента. Постоянная актуализация страниц усиливает возвращение пользователей. Со временем они осваивают навигацию, овладевают внутренними поисковыми возможностями сайта. Для эффективного продвижения библиотечных ресурсов можно рекомендовать развитие внутрисайтовой навигации на основе ручной или автоматической генерации специализированных предметно-тематических баров, связывающих контекстно близкие страницы или ресурсы со сходной метаинформацией. Развитие гипертекстовых связей между различными группами ресурсов резко усиливает показатели вовлеченности аудитории – время пребывания на сайте, глубину просмотра страниц и пр.

По данным Google Analytics самыми посещаемыми разделами сайта Отделения являются «Библиография» (/biblio/), еженедельная «Выставка новых поступлений» (/acquisitions/) и «Наука» (/science/) – с показателями соответственно 125, 96 и 86 тыс. просмотров за полтора года. В первой десятке самых популярных разделов равномерно присутствуют все типы ресурсов: и библиографические указатели, и навигаторы, и полнотекстовые материалы (см. Рис. 1). Однако их индивидуальные метрики разнятся в значительной степени. Так, например, показатель отказов для библиографических ресурсов находится в диапазоне от 83,54% до 87,94%, тогда как у веб-навигатора «SciGuide» он составляет лишь 62,49%, а у входной (домашней) страницы – 49,08%.

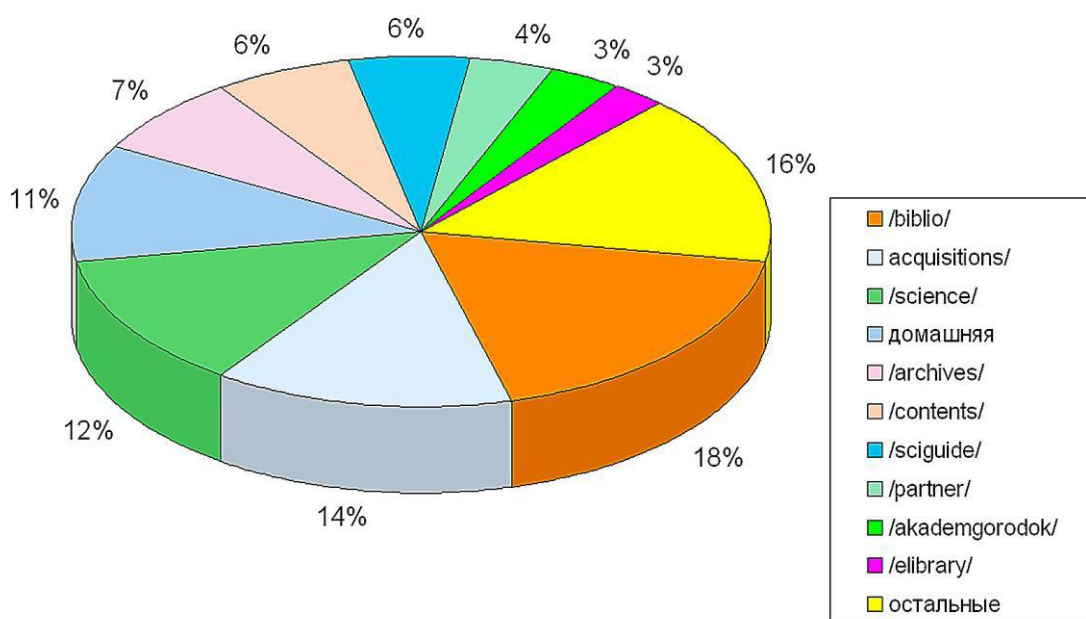


Рис. 1. Десятка самых посещаемых тематических разделов сайта www.prometeus.nsc.ru по количеству просмотров (данные Google Analytics за 01.03.2016 – 31.08.2017)

Вследствие этих расхождений индивидуальные особенности типологически разных групп библиотечных ресурсов представляли для нас особый интерес. При проведении исследования ресурсы были сегментированы на пять главных типов – *библиографические, полнотекстовые, фактографические, веб-навигаторы и смешанные (комплексные)*. По каждой группе производились подсчёты количества страниц и с помощью Google Analytics определялись ключевые показатели эффективности ресурсов (KPI = key performance indicators), такие как количество просмотров, доля отказов, глубина проникновения и пр. Для корректного сравнения разномасштабных ресурсов вычислялись коэффициенты обращений (K_o), представляющие собой отношение числа просмотров к суммарному количеству страниц в группе.

На основе этого показателя можно констатировать, что самыми популярными библиотечными ресурсами на сайте *www.prometeus.nsc.ru* являются веб-навигаторы (в среднем более 230 просмотров на каждую страницу) и фактографическая информация (справочная, энциклопедическая, контактная, новостная). Этот факт подтверждает и статистика «входов» на сайт, где однозначно лидируют две первые страницы «SciGuide» (page01.ssi и page03.ssi). На них в сумме приходится 12 тыс. просмотров. По длительности просмотров и глубине проникновения вне конкуренции находятся полнотекстовые ресурсы и библиографическая информация (см. *Таблицу 3*).

Таблица 3. Сводка ключевых показателей эффективности сайта *www.prometeus.nsc.ru*, по отдельным группам ресурсов (данные Google Analytics за 01.03.2016 – 31.08.2017)

тип ресурса	объем страниц	просмотры	коэффициент обращений K_o	уникальные просмотры	длительность просмотра	отказы	выходы
библиографические	14 339	316 170	22,0	249 740	00:01:39	82,51%	58,29%
полнотекстовые	3 924	92 208	23,5	76 372	00:01:50	78,80%	52,10%
фактографические	418	44 525	106,5	29 499	00:01:37	55,20%	31,94%
веб-навигаторы	210	48 368	230,3	32 695	00:01:05	38,40%	30,75%
комплексные	24 256	408 002	16,8	278 603	00:01:58	75,62%	49,52%
в целом по сайту	16 745	690 967	41,3	512 759	00:02:00	78,46%	54,13%

В целом, данные Google Analytics позволили довольно точно оценить размеры и качественные характеристики трафика сайта Отделения. После структуризации отдельных метрик появилась возможность вычленив наиболее продуктивную часть посещений – «эффективный трафик», связанный с обслуживанием реальных пользователей, а не ботов.

Заключение. Основные выводы и рекомендации по оптимизации ресурсов сайта Отделения и повышению эффективности их использования заключаются в следующем:

1. библиотечные сайты, распространяющие контент и имеющие информационную природу, имеют свою специфику, которая отражается в источниках трафика, метриках новых посетителей, отказов и др.;
2. высокие значения показателя отказов в отношении библиотечного контента и осо-

- бенно библиографических ресурсов не имеют определяющего значения;
3. для продвижения библиотечной информации важную роль играют показатели «проникновения» и устойчивости аудитории (возвращающихся посетителей);
 4. для улучшения этих метрик развитие библиотечного сайта должно опираться на совершенствование навигации, усиление тематических связей и контекстной близости документов, а также постоянную актуализацию ресурсов;
 5. мониторинг показателей KPI удобнее всего вести с помощью Google Analytics и Яндекс-Метрики, а системы лог-анализа (AWStats и др.) рекомендуется использовать исключительно как инструмент текущего управления сайтом.

ЛИТЕРАТУРА

- [1] Как показатель отказов влияет на конверсию // Айрико: [сайт]. 26 октября 2015 г. <https://airee.co/ru/показатель-отказов-влияет-конверсия/> (дата обращения 30.09.2017).
- [2] Кошик А. Веб-аналитика 2.0 на практике. Тонкости и лучшие методики.: пер. с англ. М.: Вильямс, 2017. 526 с.
- [3] Леонтьев А.А. Статистика посещаемости библиотечного сайта и квантовый принцип неопределённости // Научные и технические библиотеки. 2016. № 4. С. 30-34.
- [4] Скородумов П.В., Холодев А.Ю. Анализ подходов и инструментальных средств анализа статистики посещения веб-сайта научной организации // Вопросы территориального развития. 2015. № 9 (29). С. 1-13.
- [5] Farney T., McHale N. Web Analytics Strategies for Information Professionals: a LITA guide / Library & Information Technology Association. Chicago: American Library Association, 2013. 236 p.