

Заведующий отделом
Национальной библиотеки
Беларуси

БИБЛИОТЕКА В ИНТЕРНЕТ-СРЕДЕ: ЧТО И КАК СЧИТАТЬ?

Развитие информационных технологий требует от библиотек новых форм информационно-библиотечного обслуживания. Поэтому веб становится важной частью их существования. Библиотеки разрабатывают сайты, открывают свои «пространства» в социальных сетях, ведут онлайн-каталоги, создают электронные коллекции, осуществляют электронную доставку документов и справочное обслуживание.

Работая в офлайне, библиотеки разработали большое количество показателей, анализируя которые они могут оценить эффективность и успешность своей работы. Онлайн-среда – не исключение: только решения здесь необходимо принимать на основе веб-аналитики. Для начала необходимо ответить на основные вопросы: «Что считать?» и «Как считать?»

Что считать?

Какие основные статистические показатели необходимо учитывать в интернет-среде? Существует целый ряд показателей, характеризующих эффективность работы библиотеки в интернет: количество пользователей, посетителей, посещений, просмотров страниц, поисков, время пребывания на сайте, количество отказов и т. д. Как и в офлайне, так и в интернет-среде все наши сервисы направлены на удовлетворение информационных потребностей пользователей. Поэтому показатели количества пользователей, посетителей и посещений образуют основу для любой интернет-статистики. Именно их мы и рассмотрим в данной статье.

В соответствии с СТБ 7.20–2000 «Библиотечная статистика. Основные положения»: удаленный (виртуальный) пользователь – это физическое или юридическое лицо, которое зарегистрировано в базе данных пользователей на сайте библиотеки.

Стандарт ISO 2789 «Информация и документация – Международная библиотечная статистика» определяет виртуальных посетителей и посещения следующим образом:

посетитель веб-сайта – это или уникальная и идентифицированная программа веб-браузера или идентифицированный IP-адрес, который получил доступ к страницам веб-сайта библиотеки.

виртуальное посещение – это пользовательский запрос на библиотечный сайт, поступивший в би-

блиотеку не из здания библиотеки и не зависящий от количества просмотренных страниц или элементов.

Как считать?

Виртуальные пользователи

На сегодняшний день наиболее надежным и удобным способом хранения информации о пользователях является база данных (БД).

При посещении неким лицом сайта или обращении к одному из сервисов сайта, ему необходимо зарегистрироваться, то есть оставить свои индивидуальные данные (как правило имя, пароль, адрес электронной почты). С учетом того, что пользователь чаще всего обязан подтвердить свои регистрационные данные (с адреса своей электронной почты), преимущество данного способа заключается в наибольшей точности идентификации пользователей и их повторных визитов. Учет общего количества виртуальных пользователей осуществляется по количеству идентификаторов (ID) записей в БД пользователей на сервере библиотеки.

Как получить регистрационные данные от пользователя? Обобщенный алгоритм регистрации следующий:

- предоставить пользователю форму для ввода регистрационных данных;
- проверить полученные данные (заполнение обязательных полей, проверка их на корректность);
- создать учетную запись пользователя в БД (записать регистрационные данные);
- актуализировать учетную запись для аутентификации пользователя и доступа к ресурсам.

Реализация этого алгоритма на практике зависит от программной платформы вашего сайта. Для большинства систем управления содержанием сайта механизмы регистрации реализованы в виде отдельных подключаемых модулей, внедрение которых не представляет особых трудностей. В случае индивидуальной разработки сайта придется прибегнуть к помощи разработчиков или использовать готовые программные решения. Конечно, необходимо позаботиться о конфиденциальности регистрационных данных пользователя.

Возникает вопрос: как заставить регистрироваться пользователя, если многие материалы на

Информационные центры

сайте находятся в открытом доступе? Отсюда вытекает и «недостаток» данного способа идентификации: он применим, в основном, к отдельным сервисам, таким как электронный каталог, справочная служба и др. В то же время, количество зарегистрированных пользователей – показатель востребованности предлагаемых библиотекой виртуальных ресурсов. Регистрируя своих пользователей, библиотека получает очень важное преимущество: возможность взаимодействия с пользователями для совершенствования своей деятельности. Поэтому она должна вести постоянную работу по созданию уникальных материалов и сервисов, чтобы привлекать новых пользователей и «подогревать» их интерес.

Посетители и посещения

Как уже отмечено, большинство материалов сайта находится в открытом доступе и для их просмотра человеку нет необходимости проходить процедуру регистрации. Тем не менее, владельцу сайта важно знать степень востребованности таких материалов. Показатели «количество посетителей и посещений» позволяют получить эту информацию.

Между человеком, использующим виртуальный ресурс, и самим ресурсом всегда существуют посредники: компьютеры, браузеры, коммуникационные устройства и каналы связи. Как упоминалось выше, для определения пользователя необходимо использовать регистрацию. В остальных случаях мы имеем дело с посетителями. Поэтому существует два основных способа идентификации посетителей: по IP-адресу компьютера посетителя и по его веб-браузеру. Каждый из которых имеет свои преимущества и недостатки.

Идентификация по IP-адресу. IP-адрес – это уникальный адрес компьютера в сети интернет. Преимущество метода заключается в его простоте: уникальный IP – уникальный посетитель. Недостаток – это погрешность, которая состоит в том, что сайт могут посетить несколько лиц с одним и тем же IP-адресом, например, работающие через прокси-сервер, а учтены они будут как одно лицо.

Идентификация по веб-браузеру или по файлам cookies – небольшим файлам, представляющим собой данные, отправленные веб-сервером при посещении сайта библиотеки, и хранимые на компьютере лица, посетившего сайт. Механизм вычисления посетителей следующий:

1. При запросе кем-либо одной из страниц сайта инструмент веб-аналитики, используемый вами, установит в браузере данного лица уникальный cookie-файл.

2. Этот cookie-файл остается в браузере постоянно и имеет уникальный идентификатор.

3. При последующем посещении сайта сохраненный идентификатор cookie-файла используется для опознания браузера.

Преимущества идентификации по файлам cookies: статистика виртуальных посещений в большей степени сопоставима с физическими посещениями, чем при использовании метода идентификации по IP-адресу.

Недостатки:

– разные лица, использующие один браузер, учитываются как один посетитель;

– лица, использующие более одного браузера, учитываются как разные посетители;

– возможность завышенной оценки количества новых посетителей из-за отклонения или удаления cookie-файла. По разным оценкам от 10 до 30% пользователей в течение месяца теряют cookie (переустанавливают программное обеспечение, принудительно удаляют личные данные и т. д.).

Одним из важных показателей в интернете являются посещения, которые отражают тот факт, что некто посетил ваш сайт и уделил ему внимание, запрашивая страницы. На техническом языке это можно назвать *сеансом* взаимодействия с сайтом. Учет количества посещений осуществляется посредством файлов cookies.

Алгоритм измерения посещений можно упрощенно описать следующим образом:

1. При запросе кем-либо одной из страниц интернет-ресурса инструмент веб-аналитики, установленный на сайте, запускает сеанс данного лица из используемого им браузера и присваивает ему уникальный идентификатор;

2. Последующие запросы данного лица присоединяются к этому сеансу;

3. После ухода с сайта просмотренные данным лицом страницы «привязываются» к данному сеансу (посещению).

Из алгоритма понятно, что сеанс взаимодействия посетителя с сайтом может включать один и более просмотров страницы. Если посетитель оставляет браузер открытым и отходит от компьютера, сеанс принудительно прерывается. Это так называемый тайм-аут сеанса, как правило, равный 30 минутам, и который многие инструменты веб-аналитики позволяют настраивать.

Инструменты для получения статистических данных

Основными методами сбора информации о количестве посетителей (посещений) и их поведении на сайте являются анализ лог-файлов сервера и использование статистики, получаемой от внешних счетчиков.

Анализ лог-файлов сервера, на котором расположен ваш сайт

Веб-сервер протоколирует в хронологическом порядке все события в специальный журнал (англ. log) или лог-файл. В лог-файле фиксируется информация о посетителе и его действиях: IP-адрес компьютера, используемый браузер, проведенное на сайте время, просмотренные страницы и т. д. С помощью специальной программы (лог-анализатор), установленной, как правило, на веб-сервере, можно проанализировать записи лог-файла и получить сводные сведения поведения посетителей сайта за определенный период. Обработанные данные лог-анализатор хранит в своем внутреннем архиве и представляет их в понятном и удобном для человека виде. Доступ к статистике осуществляется со специальной страницы сервера.

Преимущества:

– более точное фиксирование всех посещений;
– получение специфических отчетов: учет загрузки файлов, хакерские атаки; настройка под нужды владельца сайта;

Информационные центры

- независимость от пропускной возможности канала и загруженности сервера;
- статистика сайта принадлежит владельцу сайта.
- Недостатки:
 - не совсем достоверная идентификация посетителя (например, при посещении разными лицами с одного IP-адреса);
 - наличие необходимого дискового пространства для хранения лог-файла и файла данных программы-анализатора, вероятность потери данных по причине человеческого фактора;
 - установка, настройка и обслуживание лог-анализатора требует времени и квалифицированного персонала;
 - некоторые функции реализованы только у коммерческих продуктов.
- Среди программ-анализаторов лог-файлов сервера можно отметить:
 - AWStats (<http://awstats.sourceforge.net>) (бесплатная);
 - Webalizer (<http://www.webalizer.com>) (бесплатная);
 - Weblog Expert (<http://www.weblogexpert.com/>) (30-дневная триал-версия);
 - Analog (<http://www.analog.cx>) (бесплатная);
 - AlterWind Log Analyzer Professional (<http://www.alterwind.ru>) (30-дневная триал-версия), есть бесплатная версия AlterWind Log Analyzer Lite;
 - анализатор логов из пакета Semonitor (<http://semonitor.ru>) (бесплатная демо-версия);
 - Web Trends (<http://www.webtrends.com>);
 - FastStats Analyzer (<http://www.mach5.com>) (бесплатная демо-версия);
 - Advantage Web Log Analyzer (<http://www.weblogalyzer.biz>).

Установка внешних счетчиков

Счетчик (англ. counter) – это особая программа-сценарий (как правило, JavaScript), для активизации которой требуется установка специального кода на странице сайта. При загрузке страницы сценарий отправляет данные на специализированный сервер статистики, который их агрегирует и представляет с определенным интервалом времени в виде различных отчетов.

Преимущества:

- бесплатность при базовом функционале, удовлетворяющем потребности владельца небольшого сайта;
- простой и быстрый способ установки;
- возможность более точного определения уникальных посетителей, повторных визитов;
- возможность определения настроек браузера посетителя, параметров экрана;

- высокая скорость получения данных (практически в режиме реального времени);
- отсутствие нагрузки на собственный сервер, нет необходимости технического сопровождения системы;
- возможность участия в рейтингах сайтов схожей тематики;
- возможность сравнения аудиторий сайтов, получения среза аудитории по различным параметрам;
- актуальные данные по географии посетителей.
- Недостатки:
 - недостоверность статистики посещений в случае, когда пользователь быстро покинет страницу и счетчик не успеет загрузиться (фиксируется в лог-файле сервера);
 - сервер и данные не принадлежат владельцу, поэтому возможны ограничения на период хранения данных, на предоставляемые отчеты (в зависимости от выбранного тарифа и сервиса);
 - отсутствие информации по загрузке файлов, по трафику;
 - зависимость от пропускной возможности канала и загруженности сервера статистики; постоянные обращения к сервису статистики могут замедлять работу сайта (особенно при большом количестве счетчиков).

Из наиболее известных инструментов здесь можно отметить LiveInternet, Google Analytics, Яндекс Метрика, Rambler's Top100, Рейтинг Mail.Ru, Openstat, HotLog и др. Из белорусских систем – Акавита, Tit.by, Indicator.by. Необходимо отметить, что указанные инструменты статистики различны по своему функционалу, поэтому при их выборе необходимо определить, достаточно ли вам простых метрик (LiveInternet, Рейтинг Mail.Ru, Акавита) или нужны гибкие инструменты анализа полученных данных (Google Analytics, Яндекс Метрика).

По данным инструмента «Wappalyzer» на 1 января 2014 г. лидирующую позицию в рейтинге установленных систем веб-аналитики в мире занимает Google Analytics (рис. 1).

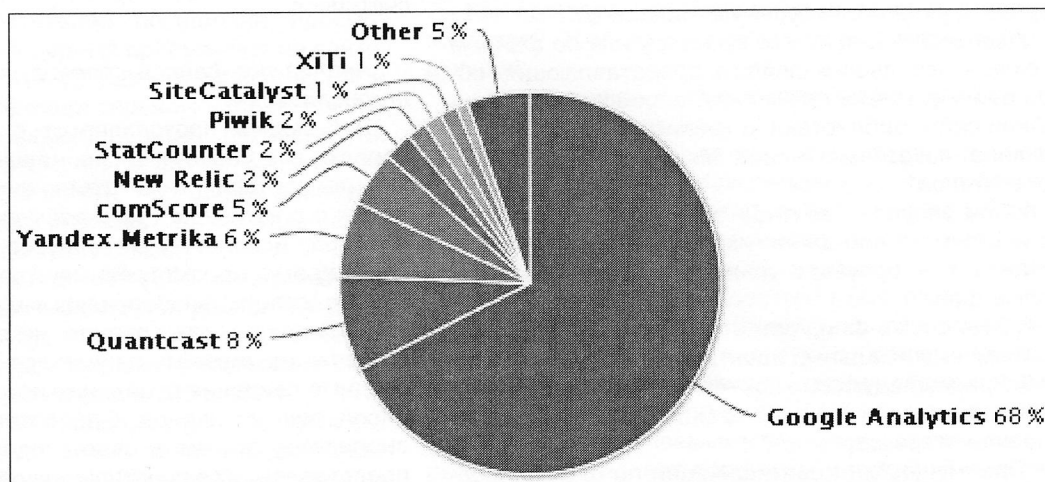


Рис. 1. Рейтинг установленных систем веб-аналитики

В рунете на 1 января 2014 г. по данным интернет-проекта «Ruward:Track» рейтинг установленных систем веб-аналитики выглядит следующим образом (рис. 2).

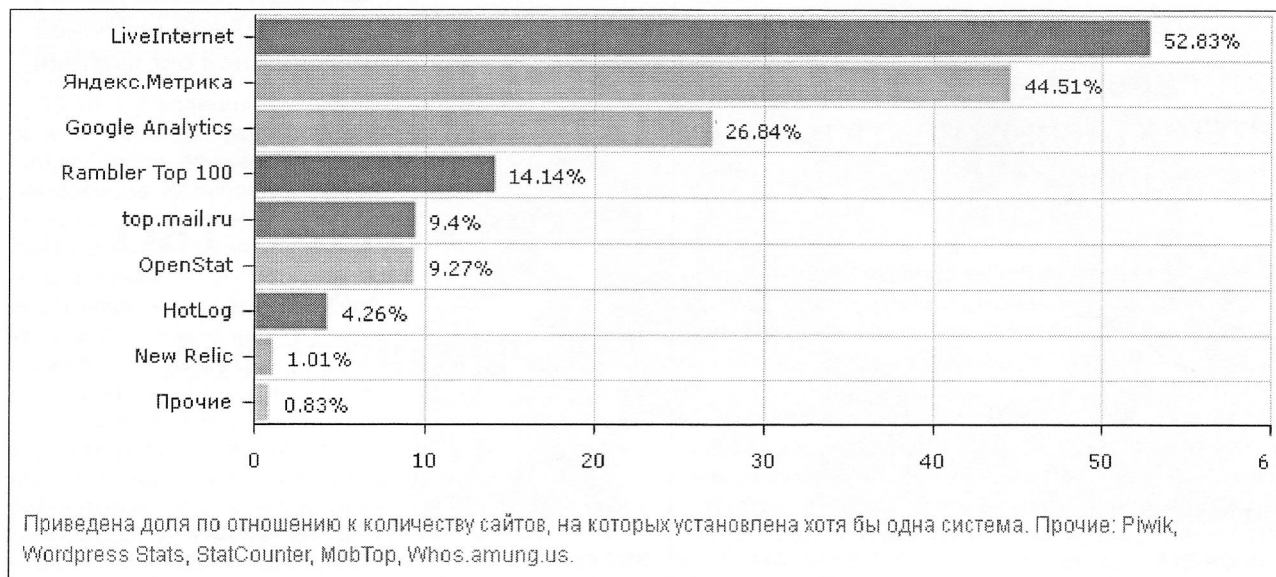


Рис. 2. Рейтинг установленных систем аналитики рунета

Какой же метод использовать?

Требования к разработке, сопровождению и эксплуатации интернет-сайтов определены в СТБ 2105–2012 «Информационные технологии. Интернет-сайты государственных органов и организаций. Требования». В стандарте указано, что как минимум одна из систем статистики интернет-сайта должна основываться на данных аудита сервера, на котором размещен сайт, то есть на анализе лог-файла сервера.

Также в п. 7.3 этого документа установлены/приведены требования к системе интернет-статистики, в соответствии с которыми она должна:

- поддерживать основные форматы файла журнала аудита сервера (Apache Log Format, W3C Extended Log File Format, IIS Log File Format);
- иметь настройки, определяющие собственный формат файла журнала аудита;
- поддерживать анализ файлов журнала аудита сервера большого объема (превышающего 100 Мб);
- поддерживать архивный формат файлов журнала аудита сервера;
- осуществлять горячий резерв файлов журнала аудита сервера;
- анализировать наличие технических проблем (ссылки на несуществующие ресурсы, перегрузка интернет-сайта);
- генерировать отчеты статистики по обращениям программного обеспечения, посетителям и объемам информации по датам с возможностью выбора интересующего периода.

Как правило, при размещении сайта на внешнем хостинге, услуга статистики по данным лог-файла сервера предоставляется провайдером бесплатно. В случае размещения сайта на собственном сервере придется устанавливать программы самостоятельно.

Учитывая преимущества и недостатки методов определения статистики использования сайта, можно сказать, что счетчики и лог-анализаторы являются хорошим дополнением друг друга. Наличие системы статистики на основе анализа лог-файла сервера является обязательным условием. Комбинация этой системы с бесплатными счетчиками позволяет

осуществить разносторонний анализ пользовательской аудитории сайта библиотеки.

ЛИТЕРАТУРА

1. **Веб-аналитика 2.0** на практике : тонкости и лучшие методики : [перевод с английского] / Авинаш Кошик. – Москва [и др.] : Диалектика : Вильямс, 2012. – 526 с.
2. **Термины** и определения – Яндекс.Помощь. Метрика [Электронный ресурс]. – Режим доступа: <http://help.yandex.ru/metrika/reports/glossary.xml>. – Дата доступа: 01.01.2014.
3. **Бібліятэчная** статыстыка. Асноўныя палажэнні = Библиотечная статистика. Основные положения: СТБ 7.20–2000. – Минск : Дзяржстандарт, Б. г. (2000). – 6 с. – (Сістэма стандартаў па інфармацыі, бібліятэчнай і выдавецкай справе = Система стандартов по информации, библиотечному и издательскому делу) (Дзяржаўны стандарт Рэспублікі Беларусь).
4. **Information and Documentation** – International Library Statistics: International Standard ISO 2789 : 2006. – Geneva, 2006. – 61 p.
5. **Счетчики** и системы веб-аналитики за январь 2014. Ruward Track – все рейтинги рунета [Электронный ресурс]. – Режим доступа: <http://track.ruward.ru/analytics>. – Дата доступа: 01.01.2014.
6. **Analytics** | Wappalyzer [Электронный ресурс]. – Режим доступа: <http://wappalyzer.com/categories/analytics>. – Дата доступа: 01.01.2014.
7. **Журнал** «КомпьютерПресс» | Обзор решений для анализа посещаемости сайта [Электронный ресурс]. – Режим доступа: <http://www.compress.ru/article.aspx?id=18178&iid=842>. – Дата доступа: 01.01.2014.
8. **Веб-аналитика**: Выбор инструментов статистики [Электронный ресурс]. – Режим доступа: <http://netology.ru/articles/1-veb-analitika-vybor-instrumentov-statistiki>. – Дата доступа: 07.01.2014.
9. **СТБ 2105–2012** – Национальный фонд ТНПА [Электронный ресурс]. – Режим доступа: <http://tnpa.by/ViewFileText.php?UrlRid=199405&UrlOnd=%D1%D2%C1%202105-2012>. – Дата доступа: 07.01.2014.

Summary

Library statistics and the application of web technologies in its formation.