

УНИКАЛЬНЫЙ ПОСЕТИТЕЛЬ БИБЛИОТЕЧНОГО САЙТА

С.К. КАНН

Государственная публичная научно-техническая

библиотека СО РАН (Новосибирск)

e-mail: serge@prometeus.nsc.ru

Аннотация.

Показатель «уникальных посетителей» приближен к реальному уровню посещаемости библиотечных сайтов и может использоваться в качестве адекватного инструмента библиотечной веб-статистики. Он гораздо меньше подвержен искажениям, характерным для показателей «доступов» и «обращений», поэтому может считаться индикатором реальной, «очеловеченной» посещаемости электронных библиотек. Проведённые тесты подтверждают, что у сайта www.prometeus.nsc.ru сформировалась устойчивая аудитория пользователей, изучение которой позволит более эффективно удовлетворять их запросы.

Уникальным посетителем (УП) называют пользователя сети, идентифицированного в качестве уникального клиента за определённый промежуток времени, чаще всего за сутки. Очевидно, это определение носит самый общий характер и далеко не идеально. Его применение в веб-аналитике сталкивается с массой сложностей. Главная проблема заключается в том, что между собой «общаются» программы и компьютеры, другими словами, именно через них осуществляется коммуникация между реальными пользователями сети. В блогосервисе blogspot.com, например, за одним IP-адресом скрывается миллион различных сайтов [1]. Наоборот, всякие «анонимайзеры», утилиты очистки служебной информации и прочие технологии, которые удаляют следы пребывания посетителей в сети, увеличивают количество IP-адресов, делая клиентов Интернета не такими уж и «уникальными».

Все эти трудности, тем не менее, не означают, что нужно подождать до лучших времён и отказаться от разработки подходов к Интернет-статистике. Сколь сложным является понятие УП, настолько же и важным для описания целевой аудитории сайтов, характеристики уровня востребованности ресурсов, степени известности («раскрученности») сайтов в Интернете. *Трастовые сайты*, имеющие высокий уровень доверия со стороны поисковых машин [2], как правило, располагают и широкой аудиторией пользователей, и большим количеством ежедневных уникальных посетителей.

На наш взгляд, показатель УП максимально приближен к реальному уровню посещаемости и может использоваться в качестве адекватного инструмента для статистических сравнений. Он в гораздо меньшей степени искажается автоматическими средствами генерации трафика – ботами, валидаторами, всякими «хоиками»-«лоиками» (HOIC-LOIC), предназначенными для DDoS-атак. По крайней мере, показатель УП проще отследить и внести необходимые уточняющие коэффициенты. Многие системы статистики приспособлены отсекают автоматические средства индексации с помощью специально назначенных временных интервалов для доступов, осуществляемых в пределах 15–30 секунд. Непосредственно с показателями «обращений» это сделать гораздо труднее, так как многие роботы не представляются серверам, другими словами, не отмечаются в файле *robots.txt* и не сообщают необходимую служебную информацию в поле юзер-агента лог-файла. Таким образом, с некоторыми правками показатель УП может расцениваться как индикатор реальной, «очеловеченной» посещаемости электронных библиотек.

Наиболее часто используется метод идентификации пользователей по IP-адресу, но возможны и другие способы, использующие пересылку «куки» (*cookies*) или регистрацию пользователей по логину-паролю (*identified users*). Автоматический учёт всех IP-адресов представляется незаметным и наименее обременительным средством статистики, не отвле-

кающим посетителей сайта от непосредственных задач, как в случае принудительного заполнения анкет и «залогинивания». Проигрывая в информативности, учёт на основе лог-файла выигрывает «на оборотах» – в режиме свободного и массового доступа пользователей.

До последнего времени увеличение количества УП в значительной степени поддерживалось простым («физическим») развитием российского сегмента сети, происходившим на протяжении всех 2000-х годов. Количество УП росло за счёт подключения новых офисов и квартир, сооружения современных базовых станций, прокладки оптоволоконных линий и мобилизации Интернета. В расширявшемся телекоммуникационном пространстве действовали противоположные тенденции. С одной стороны, непрерывно увеличивалось количество клиентов, то есть непосредственных потребителей информации. Они спешили освоиться в сети, быстро совершенствовали навыки поиска и навигации. Вместе с тем, не менее стремительными темпами росло число новых сайтов и ресурсов, «растаскивавших» этих пользователей, предлагая им широкий выбор конкурентных продуктов и услуг.

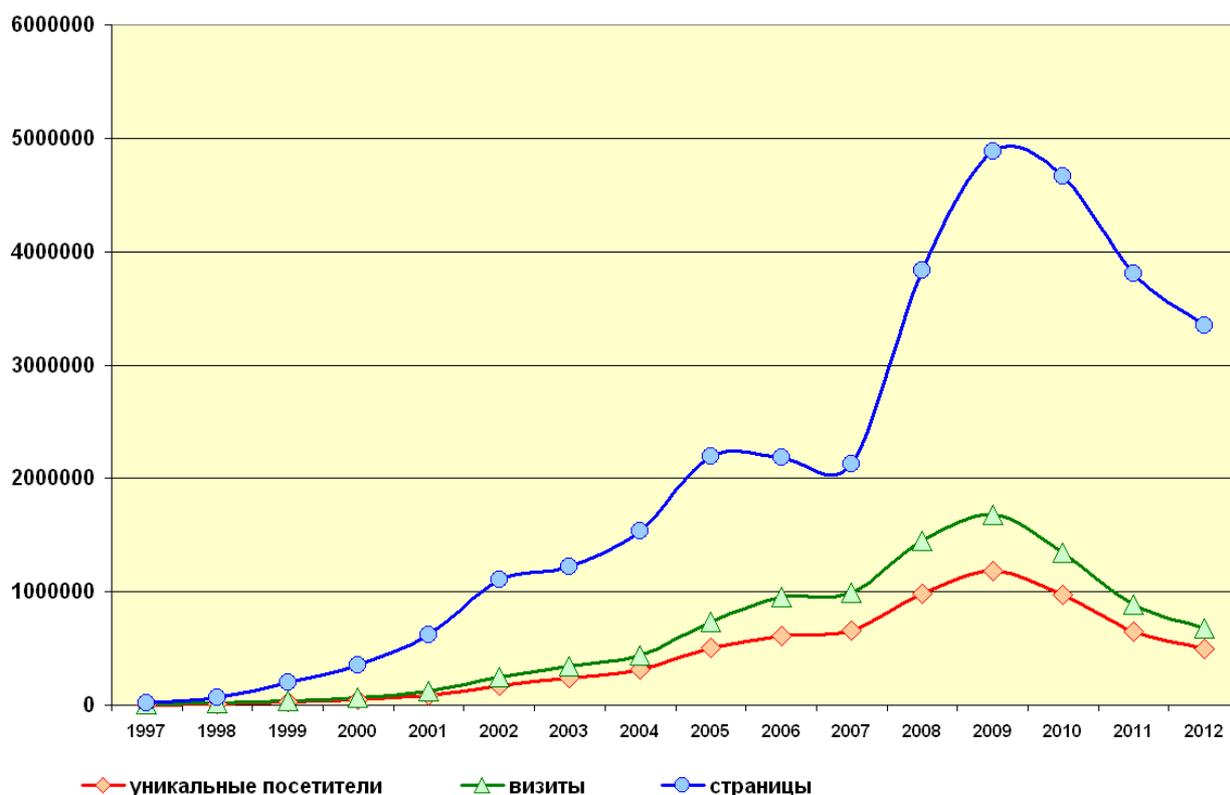


Рис. 1. Основные показатели сайта www.prometeus.nsc.ru за 1997–2012 гг. (по данным статистической системы [AWStats](#) 6.5)

Показатели библиотечного сайта отделения ГПНТБ СО РАН www.prometeus.nsc.ru непрерывно увеличивались на протяжении 12 лет и особенно резко в 2007–2009 гг. При этом показатель уникальных посетителей нарастал более плавно и устойчиво, что свидетельствовало в пользу формирования стабильной аудитории пользователей и наборе некоторого «веса» и популярности в библиотечном Интернет-сообществе (рис. 1). Статистические показатели сайта в 2009 г. были так велики, что серверная программа анализа лог-файлов [AWStats](#) не справилась с формированием годового отчёта и его пришлось досчитывать помесячно. В мае 2009 г. сайт библиотеки принял 138 тыс. уникальных посетителей, а количество обращений в январе и ноябре того же года переваливало за полмиллиона.

После 2009 года начался спад посещений, вызванный, по-видимому, целым рядом причин – социально-экономических, демографических и чисто библиотечных. Среди основных библиотечных проблем можно назвать следующие:

- 1) Традиционные библиотеки стали испытывать трудности перехода к новой парадигме своей деятельности и к новым формам обслуживания.
- 2) Ускоренное развитие новых технологий, изменчивость ресурсов и мобильность пользователей создали виртуальную среду, противопоставленную традиционной и весьма консервативной деятельности прежних библиотек.
- 3) Уход основной массы читателей в Интернет и переход преимущественно к «электронному чтению» обусловили резкое снижение реальной посещаемости.
- 4) Немногие сайты библиотек готовы дать своим читателям новые информационные продукты и услуги, а оцифровка значительной части актуальных и спрашиваемых традиционных ресурсов наталкивается на закон об авторском праве.
- 5) Слабое знание целевой аудитории сайтов и поведения пользователей в электронной среде, недостоверная веб-статистика, отсутствие методик выявления предпочтений посетителей ограничивает эффективность обслуживания в новых условиях.

Трансформация библиотечной деятельности фокусируется преимущественно на новых коммуникационных и маркетинговых возможностях – создании собственных «кабинетов» в Twitter, Facebook и Youtube, на проведении вебинаров и пр., – но не на первоочередном развитии информационной и методологической компоненты. Между тем, будущее научных библиотек, о котором сейчас много говорится и пишется, связано с усилением именно *информационных* и *методологических* функций, с переходом к специализированным библиотечно-информационным центрам, генерирующим новый контент как по форме, так и по содержанию, принципиально новые ресурсы навигационного, справочного, библиографического, реферативно-аналитического типа. Нужно развивать информационные и библиографические навыки пользователей электронных библиотек, обучать их работе с новыми компьютерными продуктами, базами данных и пр. В связи с этим от библиотекарей, работающих в электронной среде, требуется хорошее знание своей целевой аудитории, для которой следует *целенаправленно* создавать необходимые ресурсы.

На протяжении полутора десятилетий сайт отделения ГПНТБ СО РАН ведёт статистический мониторинг доступов, используя аналитическую систему [AWStats](#), а также системы «баннерной» статистики, предполагающие внедрение особого кода на всех страницах с последующей передачей служебной информации на центральные сервера проектов [Rambler Top100](#) и «Актуальные ресурсы NSC.Ru». Второй проект, можно сказать, «благополучно

скончался» в этом году, поэтому 31 октября сайт [www.prometeus.nsc.ru](#) вступил в новый рейтинг [LiveInternet](#) [3]. Уже через сутки он поднялся на четыре тысячи мест и обосновался примерно на 200-х местах в таблице учтённых ресурсов [LiveInternet](#) *научного профиля*.

Достоинства новой системы статистического мониторинга, на наш взгляд, заключаются в широком представлении параметров, касающихся



Рис. 2. Половозрастная структура посетителей [www.prometeus.nsc.ru](#) по данным «Яндекс.Метрики» за период с ноября 2011 по ноябрь 2012 г.

именно уникальных посетителей сайта. [LiveInternet](#) позволяет получить срезы, описывающие размер пользовательской аудитории, включая её активное «ядро», тех, кто регулярно посещает сайт каждую неделю или месяц, провести сравнение аудиторий, получить данные о количестве сессий на каждого посетителя, о возвратах, просмотрах на одного посетителя, просмотрах за сессию, длительности сессий и т. п. Эти данные значительно дополняют другие

средства статистики, используемые при анализе функционирования вебсайта отделения ГПНТБ СО РАН.

Одним из эффективных инструментов, установленных на сайте весной 2011 г., является [«Яндекс.Метрика»](#), также позволяющая лучше понять, с каким пользователем мы имеем дело. В частности, благодаря алгоритмам собственной разработки [4], [«Яндекс.Метрика»](#) предоставляет данные о половозрастной структуре наших посетителей (рис. 2–3). Из этих данных видно, что существенную часть составляют студенты и учащиеся в возрасте от 18 до 24 лет, а люди в самом активном и трудоспособном возрасте от 35 до 45 лет занимают наименьшую долю. Очевидно, что здесь есть повод задуматься, какого рода ресурсы и услуги могли бы привлечь их внимание. Любопытно, что сайт в достаточной степени интересен молодым девушкам (до 18 лет) и мужчинам от 45 лет и старше. На наш взгляд, это свидетельствует об универсальном, многостороннем и гуманитарном характере большинства представленных ресурсов.

На рис. 3 заметно, что по сезонным особенностям посещаемости вебсайт библиотеки практически ничем не отличается от традиционных читальных залов. С января по июль доступ пользователей, условно говоря, «студенческого» возраста (**оранжевая кривая**) постоянно снижается и это становится особенно заметным в начале лета. Напротив, посещаемость представителей «аспирантского» возраста (**синяя кривая**) и «возрастных» читателей, возможно, научных сотрудников (**жёлтая кривая**) достигает максимума именно летом. Посещаемость сайта в «цветущем» возрасте 35–44 лет (**зелёная кривая**) находится на стабильно низком уровне 3–5%. В начале нового учебного года (сентябрь) нарастает доступ студентов и учащихся (**оранжевая, красная** линии), а **синяя** и **жёлтая** кривые идут на спад.

Таким образом, можно уверенно говорить о том, что, в целом, библиотека вполне справляется с научно-образовательными задачами, которые перед ней поставлены. Но этот вывод носит очень общий характер. Значительной детализации требует состав пользовательских предпочтений посетителей, «цепочка» их перемещений по сайту, а также особенности конверсии, то есть определённых целевых действий, на которые рассчитывают создатели ресурсов. Опыт показывает, что сайту научной библиотеки вряд ли удастся охватить в полной мере всю множественную аудиторию пользователей глобальной сети. «Научно-библиотечная экспансия» подобного рода имеет ряд ощутимых социальных ограничений. Чтобы далеко не ходить за примером, назовём статистику телеканала «Культура», который, по данным TNS Gallup Media, даже в Москве регулярно смотрят не более 3% телезрителей, а это, как минимум, в 6 раз меньше аудитории каналов «Первый», «Россия» или НТВ [5].

С нашей точки зрения, не отвергая сложившейся массовой аудитории сайта (в значительной степени московско-петербургской – см. рис. 4), необходимо сосредоточиться на

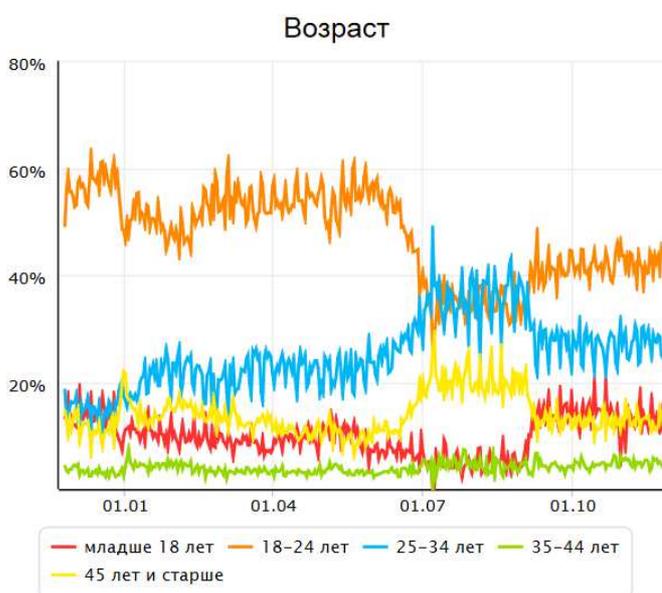


Рис. 3. Сезонная посещаемость сайта www.prometeus.nsc.ru с распределением по возрасту с ноября 2011 по ноябрь 2012 г.

«целевом ядре» пользователей – сотрудниках Сибирского отделения Академии наук, а также на тех из них, кто посещает сайт регулярно, как минимум один-два раза в неделю, кто заинтересован в наших новостях, выставках и обновлении ресурсов.

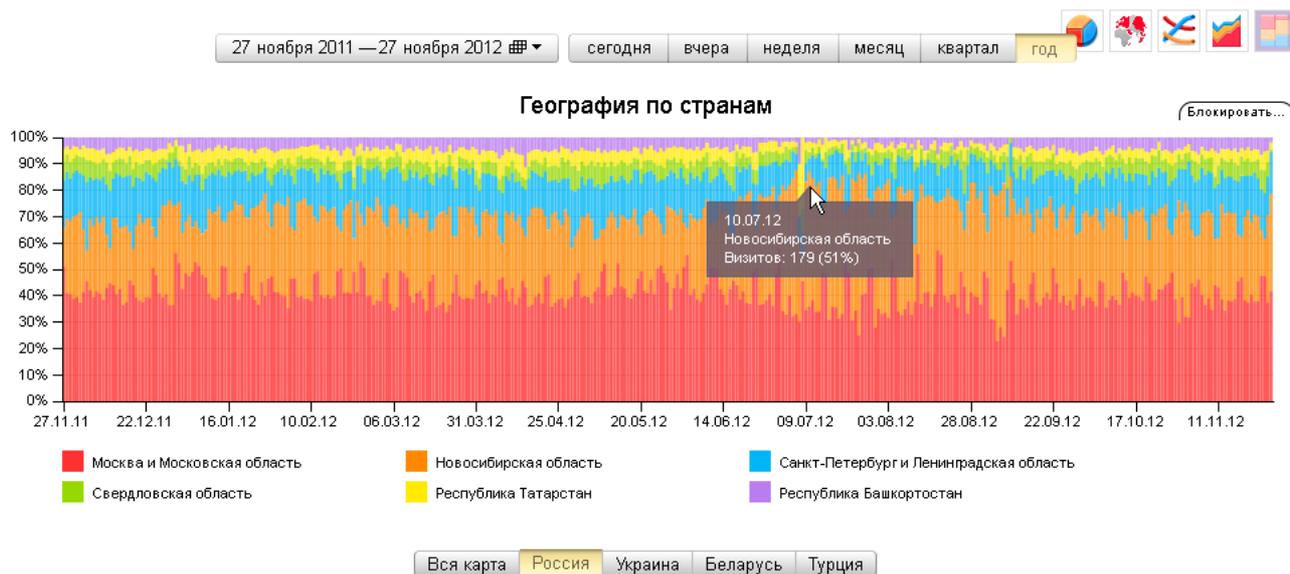


Рис. 4. География посетителей сайта www.prometeus.nsc.ru по регионам России в 2012 г. (по данным «Яндекс.Метрики»)

Генерация нового контента, в первую очередь, должна ориентироваться на их запросы, но они, безусловно, требуют постоянного и тщательного изучения. В качестве частной задачи можно попытаться если не выправить, то хотя бы приостановить развитие негативных тенденций 2010–2012 гг., представленных на рис. 1.

С самого начала своей деятельности сайт отделения ГПНТБ СО РАН отличался тем, что старался охватить максимум вероятных пользователей своей информацией, «втянуть» их в орбиту своих научно-образовательных задач. Соответствующим образом подбирались и формы для реализации данного проекта – простые, общедоступные решения, не ограничивавшие пользователей ни по операционным системам, ни по разрешению экрана, ни по браузерам. Ставка на «прозрачность» для поисковых машин, простоту кода и интуитивно понятную навигацию привели к формированию широкой, но постоянно обновляющейся аудитории потребителей информации, типичной для сообщества учащихся и студентов. Для них сайт стал удобным рабочим инструментом, а мотивацию этой аудитории прекрасно отражает статистика. По данным «Яндекс.Метрики», посетители www.prometeus.nsc.ru в 2012 году на 99,5% – это «новые посетители», то есть «уникальные пользователи, активность которых включала их самый первый за всю историю накопления данных визит на сайт». Уровень отказов по www.prometeus.nsc.ru достигает 80,3%. Иначе говоря, именно такая доля пользователей покидает сайт после всего лишь одной посещённой страницы, хотя это вовсе не говорит о плохом качестве ресурсов, а только о большой точности «информационных попаданий». Среднее время, проведённое на сайте каждым посетителем составляет 1 мин. 30 сек.

Для изучения вопроса, в какой степени ресурсы www.prometeus.nsc.ru удовлетворяют **сотрудников Сибирского отделения РАН**, мы провели несколько исследований, основанных на фильтрации данных из лог-файла. В течение 2012 г. изучались доступы к сайту с IP-адресов из трёх сеток NSCNET (диапазон адресов 194.226.160.0–194.226.191.255 и 212.192.160.0–212.192.191.255), определённых с помощью сервиса WhoIs [6]. В целом, это более 16 тыс. компьютеров. Результаты подсчётов оказались удручающими. За девять месяцев 2012 г. (по сентябрь включительно) на IP-адреса NSCNET пришлось только 11.638

РЕСУРС (url)	Количество доступов
Главная	657
Академгородок	1063
О библиотеке	237
Подразделения	155
Сотрудники (с трудами)	204
Новости	173
Архивы	363
ВНП	234
Услуги + виртуальная справка	42
Ресурсы	247
Электронный каталог -Jisls	293
Поиск по сайту	79
Библиография	99
Оглавления	80
Патентные ресурсы	176
Ресурсы МБА	218
Зал ВМИ	54
Доступ к зарубежным журналам on-line	648
Указатель отечественных журналов	299
Электронная библиотека	715
Научные школы СО РАН	635
Календарь памятных дат СО РАН	397
В.А. Коптюг	84
Н.Н. Яненко	59
SciGuide	79
Другие ресурсы о науке (дайджест и пр.)	185
English	31
Партнеры - А.П. Зарубин	144
Инфоляция	65
О вебсайте	20
Остальные обращения	3903
ВСЕГО по трём сеткам NSCNET (1-3)	11638

Рис. 5. Количество доступов из сетей NSCNET к ресурсам сайта за янв. – сент. 2012 г. (по данным лог-файла), в виде таблицы

Как впоследствии выяснилось, в самый момент нашей работы Сибирское отделение вело массовый переход со старых IP-адресов на новые, и это, вероятнее всего, отразилось на итоговых цифрах исследования. С новой информацией в руках в начале декабря был сделан повторный пересчёт – уже по сеткам всего Сибирского отделения РАН (для диапазона 84.237.0.0–84.237.128.255, то есть примерно для 33 тыс. компьютеров). Маячила надежда, что картина изменится, и, действительно, пересчёт показал другие цифры. Впрочем, и они оптимизма не внушали. За 11 месяцев 2012 г. лог-файл зафиксировал порядка 3,3 млн обращений к сайту, из которых на долю IP-адресов СО РАН, за вычетом доступа сотрудников библиотеки, пришлось лишь 29.992 или 0,9%. Если принять во внимание, что количество IP-адресов, взятых для расчёта, возросло более чем вдвое по сравнению с сентябрьским исследованием, а учётный период расширился на два месяца (октябрь-ноябрь), то мы, фактически, получили ту же самую пропорцию (0,4% по первому исследованию и 0,9% – по второму).

Таким образом, можно констатировать, что при всей сложности изучаемых объектов, веб-аналитика позволяет выявить сущностные черты и важные закономерности функционирования веб-пространства. Открытый характер этого пространства, наличие большого количества связей и активных факторов, влияющих на информационные взаимодействия, не все-

доступов из общего количества 2.810.864, то есть крохотные 0,4%. Несмотря на то, что сам характер обращений не мог не радовать, так как большинство из них направлялось к ресурсам академического профиля (в разделы «Наука», «Мемориальные библиотеки» и другие – см. рис. 5–6), однако слишком малое абсолютное количество обращений насторожило.

Примерно такой же скромный уровень использования показал подсчёт доступов к веб-навигатору научных ресурсов открытого доступа [«SciGuide»](#), который развивается совместно с отделом комплектования иностранной литературой ГПНТБ СО РАН с января 2010 г. Доступы с IP-адресов компьютеров, прописанных в СО РАН для работы с зарубежными базами данных (как правило, это машины из библиотек НИУ Новосибирска, Красноярска, Иркутска и других центров), ограничились несколькими десятками на каждую, а в 2012 г. таких доступов практически вообще не было. Сообщение по итогам исследования делалось 27.09.2012 г. на межрегиональной научно-практической конференции [«Научные библиотеки России: взгляд в будущее»](#).

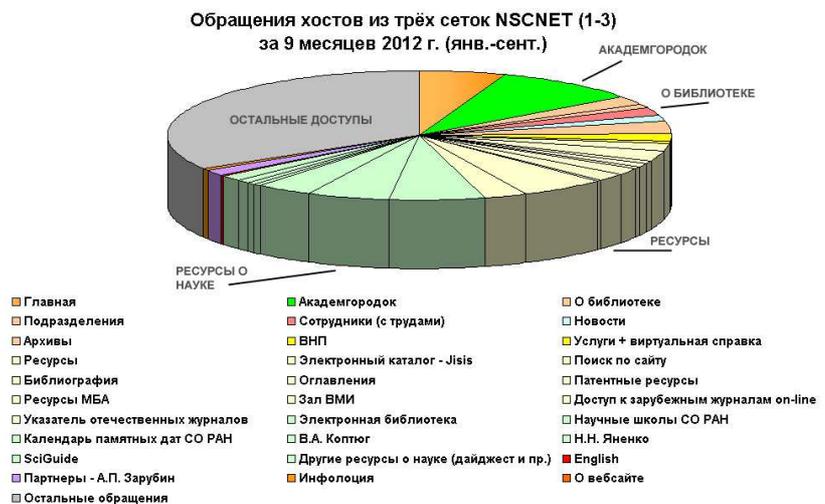


Рис. 6. Количество доступов из сетей NSCNET к ресурсам сайта за янв. – сент. 2012 г. (по данным лог-файла), в виде диаграммы

гда приводят создателей сайтов к желаемым результатам. Тем не менее, изучение процессов веб-коммуникации способствует рационализации процессов подготовки ресурсов и сознательному формированию информационного образа сайтов в нужном направлении. Задача состоит в том, чтобы оценить уникальных посетителей электронных библиотек не только с количественной, но и с качественной стороны, выявить целевую аудиторию сайта, «ядро» его регулярных пользователей, а затем целенаправленно работать на их интересы.

ЛИТЕРАТУРА

- [1]. Новая газета. 2012. 2 ноября, № 125. – URL: <http://www.novayagazeta.ru/columns/55230.html> (дата обращения: 10.12.2012).
- [2]. Понятие *трастовых* сайтов введено в работе «Борьба с веб-спамом посредством TrustRank» (2004) сотрудниками Yahoo! и Стэнфордского университета. – URL: <http://seopult.ru/library/Трасм> (дата обращения: 10.12.2012).
- [3]. Статистика посещаемости носит открытый характер и доступна по адресу: <http://www.liveinternet.ru/stat/prometeus.nsc.ru/>.
- [4]. Поиск в интернете: как ищут мужчины и женщины. – URL: http://company.yandex.ru/researches/reports/ya_on_men_women_11.xml (дата обращения: 10.12.2012); Яндекс исследовал поисковое поведение женщин и мужчин. – URL: <http://actualweb.ru/2011/06/28/Поиск-в-интернете-как-ищут-мужчины-и-же/> (дата обращения: 10.12.2012); Как Яндекс определяет пол пользователей. – URL: <http://www.cy-pr.com/news/se/3183/> (дата обращения: 10.12.2012); Как Яндекс.Метрика определяет демографию. – URL: <http://www.seocafe.info/yandex/17522-kak-yandeks-metrika-opredelyaet-demografiyu.html> (дата обращения: 10.12.2012); и др.
- [5]. Эхо Москвы. Теленеделя с 19 по 25 октября. Комментарий Арины Бородиной. 2009. 31 октября. – URL: http://www.echo.msk.ru/blog/p_o/630982-echo/ (дата обращения: 10.12.2012); Эхо Москвы. Человек из телевизора. Интервью Александра Мельмана. 2012. 4 августа. – URL: <http://echo.msk.ru/programs/personstv/914628-echo/> (дата обращения: 10.12.2012); Эхо Москвы. Человек из телевизора. Ирина Петровская – Ксения Ларина. 2012. 3 ноября. – URL: <http://www.echo.msk.ru/programs/personstv/947157-echo/> (дата обращения: 10.12.2012); и др.
- [6]. RIPE Network Coordination Centre – Database Query. – URL: <https://apps.db.ripe.net/search/query.html>.