

АНАЛИЗ ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ САЙТА «НАУЧНОЕ НАСЛЕДИЕ РОССИИ»

*Кириллов С. А., Забровская И. В., Кондратьева Е. А., Пругло О. А.
(МСЦ РАН)*

Анализ пользовательской активности тематического электронного каталога сайта «Научное наследие России»

Электронная библиотека «Научное наследие России» (<http://e-heritage.ru>) занимает важное место в ряду аналогичных проектов российского информационного пространства. Это обусловлено уникальными фондами библиотек, архивов и других научных учреждений, участвующих в проекте. Наш электронный ресурс, основанный на современных информационных технологиях, позволяет представить наиболее известные мировому сообществу фундаментальные и прикладные труды отечественных ученых, ценные коллекции, архивные материалы, а также целый ряд уникальных, редких и малодоступных изданий.

Развитая инфраструктура и техническая база организаций-участников проекта, располагающих современной системой широкополосного доступа в Интернет, а также высокий научный и технической потенциал сотрудников позволяют проводить разработку и реализацию проекта на самом высоком современном уровне.

Анализ эффективности работы тематического электронного каталога. Google Analytics

Создавая электронную библиотеку «Научное наследие России», мы интересуемся эффективностью ее работы. Одним из показателей эффективности является пользовательская активность. Нас интересует пользователи каких стран, регионов, районов, городов запрашивают информацию с сайта, какая языковая аудитория просматривает сайт. Также нас интересуют операционные системы и браузеры, которые установлены у наших пользователей. Важен анализ возрастной и социальной категории. И, наконец, мы исследуем, каким образом пользователь находит наш сайт.

Все эти исследования стали возможны благодаря тому, что в 2010 г. сайт «Научное наследие России» подключился к системе анализа основных параметров сайта Google Analytics. Google Analytics это бесплатный сервис от Google, который позволяет подсчитывать количество посещений сайта, следить за перемещениями по сайту

и вычислять количество реальных посетителей, которые пользуются сайтом, исключая посещения роботов-поисковиков и случайные запросы. С помощью Google Analytics мы можем получить разнообразную информацию о посетителях сайта. Например, определить браузер и операционную систему, которую используют посетители; узнать о разрешении экранов и типах мобильных устройств; выяснить, из какого государства и города совершается обращение к сайту, и многие другие данные.

Центральный веб-портал электронной библиотеки «Научное наследие России» представляет собой технологический комплекс, состоящий из ряда подсистем. Одной из таких подсистем является служба тематического электронного каталога

Тематический электронный каталог (ТЭК) представляет собой иерархический классификатор областей знаний на основе ГРНТИ, с помощью которого читатели могут найти интересующие их издания по отдельным наукам.

География посещений

Рассмотрим географию посещений ТЭК сайта. В период с 1 января 2012 по 21 мая 2014 г. к ТЭК сайта «Научное наследие России» обращались пользователи из 115 стран мира.

Основной массив посетителей Google Analytics относит к региону Восточной Европы (Eastern Europe), который включает всю Россию, Украину, Беларусь и Молдову. Он составил 157 404 посещений ТЭК сайта или 89,21 % от общего числа посещений.

Представленная ниже табл. 1 дает более подробную информацию о том, пользователи из каких стран чаще обращаются к ТЭК сайта (приведены только первые 10 стран).

Таблица 1. Статистика обращений к ТЭК по странам

Россия	118 731 (67,30 %)
Украина	30 603 (17,35 %)
Беларусь	4479 (2,54 %)
Страна не определена	3563 (2,02 %)
Казахстан	2669 (1,51 %)
США	1916 (1,09 %)
Германия	1687 (0,96 %)
Израиль	1328 (0,75 %)
Польша	1058 (0,60 %)
Болгария	983 (0,56 %)

Как видим, наибольший интерес наша библиотека представляет для русскоязычных пользователей стран СНГ.

Это подтверждает другой параметр Google Analytics, указывая, что у 88,11 % пользователей установлена русская раскладка клавиатуры. Стоит отметить большое количество посещений из Украины: 30 603 пользователей или 17,35 % от общего числа посещений. (Интересно, что при этом украинская раскладка клавиатуры была установлена только у 80 пользователей или 0,05 % от общего числа посещений.)

Таблица 2. Распределение пользователей по городам

1	Москва	36 068 (20,44 %)
2	Санкт-Петербург	14 876 (8,43 %)
3	Киев	12 699 (7,20 %)
4	Город не установлен	6153 (3,49 %)
5	Нижний Новгород	3580 (2,03 %)
6	Екатеринбург	3175 (1,80 %)
7	Новосибирск	3044 (1,73 %)
8	Казань	2717 (1,54 %)
9	Харьков	2685 (1,52 %)
10	Самара	2563 (1,45 %)
11	Минск	2259 (1,28 %)
12	Ростов-на-Дону	2087 (1,18 %)
13	Томск	1965 (1,11 %)
14	Львов	1720 (0,97 %)
15	Донецк	1603 (0,91 %)
16	Краснодар	1526 (0,86 %)
17	Воронеж	1526 (0,86 %)
18	Омск	1497 (0,85 %)
19	Пермь	1478 (0,84 %)
20	Днепропетровск	1478 (0,84 %)
21	Саратов	1473 (0,83 %)
22	Алматы	1295 (0,73 %)
23	Красноярск	1241 (0,70 %)
24	Одесса	1240 (0,70 %)
25	Волгоград	1215 (0,69 %)
26	Уфа	1171 (0,66 %)
27	Чернигов	1077 (0,61 %)
28	Барнаул	1047 (0,59 %)
29	Калининград	1031 (0,58 %)
30	Челябинск	981 (0,56 %)

У 11,89 % посетителей установлены различные варианты международных раскладок клавиатуры, что, примерно, соответствует международному сегменту нашей аудитории. Полученные данные указывают на необходимость создания англоязычной версии для расширения пользовательской аудитории.

Распределение пользователей по городам представлено в табл. 2. В период с 1 января 2012 по 21 мая 2014 г. на ТЭК сайта «Научное наследие России» побывали пользователи из 2206 населенных пунктов (городов) мира. В таблице представлены 30 первых городов.

Проанализировав таблицу, можно выделить Москву, С.-Петербург и Киев как наиболее «активные» города с точки зрения посещений нашего сайта. Распределение посещений по другим городам достаточно предсказуемо. Прежде всего это города-миллионники стран СНГ, в которых расположено большинство культурных, научных и учебных учреждений. Затем следуют областные центры стран СНГ и столицы стран дальнего зарубежья. Далее идут более тысячи городов с небольшим количеством жителей которые дают не более 20 % посетителей сайта. Можно предполагать, что с развитием сети Интернет в регионах увеличится количество посетителей, но уже сейчас только один перечень более чем из двух тысяч небольших городов показывает, насколько необходим и популярен наш сайт.

Количество посещений тематического электронного каталога

В табл. 3 представлено число посещений тематического электронного каталога сайта за четыре последних года с уникальных ip адресов.

В среднем, ТЭК сайта вышел на уровень 70–75 тысяч уникальных посещений в год. Проследив соотношение новых посетителей и постоянных посетителей к общему числу посетителей, можно говорить о поведении пользователя. Поведение пользователя отражает (в некоторой мере) характер аудитории нашего сайта. Поскольку специфика нашего сайта – это электронная библиотека, предоставляющая информацию, которая может быть интересна достаточно часто, то у сайта должно быть как можно больше постоянных

Таблица 3. Количество посещений ТЭК за 2011–2014 гг.

2011	2012	2013	2014 1 янв. – 20 мая
30 320	75 205	72 432	30 408

посетителей. Данный показатель мы отслеживали в течение продолжительного периода времени и он постоянно увеличивается. На 21 мая 2014 г. 40,8 % пользователей являлись постоянными посетителями ТЭК нашей Библиотеки.

Программное и аппаратное обеспечение пользователей

Теперь перейдём к анализу операционных систем и браузеров, которые установлены у наших пользователей.

Следующая табл. 4 показывает, какие браузеры используются при посещении ТЭК сайта в период с 1 января 2012 по 21 мая 2014 г. Эта информация необходима программистам для проверки корректности отображения сайта в основных браузерах.

Распределение посетителей по используемым операционным системам в период с 1 января 2012 по 21 мая 2014 г. следующее (табл. 5):

Таблица 4. Распределение посетителей по используемым браузерам

Chrome	52 703 (29,87 %)
Firefox	49 170 (27,87 %)
Opera	41 308 (23,41 %)
Internet Explorer	23 002 (13,04 %)
Safari	4837 (2,74 %)
YaBrowser	1540 (0,87 %)
Android Browser	1427 (0,81 %)
Opera Mini	1194 (0,68 %)
SeaMonkey	294 (0,17 %)
Maxthon	177 (0,10 %)

Таблица 5. Распределение посетителей по используемым операционным системам

Версия ОС	Посетителей, %
Версии Windows	91,36
Linux	2,92
Macintosh	1,71
Различные версии мобильных OS	3,41
Не определено	0,60

Таблица 6. Распределение посетителей по типу используемого устройства, %

Тип устройства/Год	2012	2013	2014
Настольный компьютер	98,54	95,72	93,03
Планшет	0,79	2,24	3,62
Телефон	0,67	2,04	3,35

Используемые операционные системы показывают, что наиболее часто пользователи заходили на сайт с персонального компьютера (стационарный или ноутбук). Однако, если мы рассмотрим этот процесс в динамике, год за годом, то можно увидеть серьезные изменения в аппаратном оснащении наших посетителей.

Из следующей табл. 6 видно, что число наших пользователей, использующих планшет и телефон, растёт очень быстрыми темпами и ежегодно увеличивается почти в два раза. Этот факт необходимо учитывать нашим программистам при изменении интерфейса электронной библиотеки «Научное наследие России».

Возраст. Социальные группы

Рассмотрим возрастную категорию пользователей ТЭК сайта (табл. 7).

Таблица 7. Возрастное распределение пользователей ТЭК

Возраст	%
18–24	27.50
25–34	33.50
35–44	15,50
45–54	12,50
55–64	5,50
65+	5.50

Ресурсы электронной библиотеки «Научное наследие России» не содержат развлекательной литературы. Представленные в Библиотеке фундаментальные и прикладные труды ученых ориентированы главным образом на учащихся, исследовательскую аудиторию, ученых и специалистов. Поэтому можно предположить, что не менее 27,50 % посетителей (в возрасте от 18 до 24 лет) сайта являются студентами, изучающими различные дисциплины. Но основной аудиторией нашей библиотеки в настоящее время

(72,5 %) являются посетители возрастом старше 25 лет – это специалисты в области истории науки, сотрудники вузов и другие лица, интересующиеся проблемами науки. Как видим, прогноз, который делался в период основания библиотеки, о том, что примерно 30 % целевой аудитории это учащаяся молодежь, но основной целевой аудиторией являются научные сотрудники и специалисты, практически оправдался.

Тематический электронный каталог. Анализ посещения

Обратимся к анализу разделов, наиболее популярных среди пользователей (табл. 8).

Таблица 8. Обращаемость пользователей по областям знаний

№ п/п	Область знаний	Количество обращений, %
1	История. Исторические науки	7
2	Философия	5
3	История атеизма, религий и церквей	5
4	История отдельных стран	4
5	Всеобщая история	4
6	История литературы	4
7	Религия. Атеизм	4
8	Государство и право. Юридические науки	4
9	Языки мира	3
10	Математика	3
11	Общее языкознание	3
12	Экономика. Экономические науки	3
13	Отдельные религии, церкви и культы	3
14	Философия и религия	3
15	Зоология	3
16	Биология	3
17	Физика	3
18	Литература. Литературоведение	3
19	Общефилософские проблемы	3
20	Химия	3
21	Геология	2
22	Языкознание	2
23	Общие вопросы физики	2

Представленная таблица показывает, какие разделы тематического каталога наиболее интересны нашим посетителям. Самыми востребованными областями науки стали разделы гуманитарных и общественных наук: истории, философии, истории религии, права и лингвистики. Затем следуют математика, экономика, биология, физика и химия. Следует отметить, что все без исключения 65 разделов тематического каталога нашей библиотеки активно используются читателями.

Рефералы

Качество контента представленного в нашей библиотеке косвенно подтверждается нашими рефералами, т.е. сайтами, которые ссылаются на наш ресурс. Мы анализируем подобные сайты и выборочно проверяем контекст ссылок на наш сайт. Наша цель – прирост ссылок на наш сайт с качественных ресурсов.

Первое место или 25,30 % всех переходов на наш ресурс дает сайт Википедия (<http://wikipedia.org>), второе место нам дает сайт БЕН РАН (<http://benran.ru>) 562 (13,18 %), на третьем месте Библиотека Российской академии наук (<http://rasl.ru>) 485 (11,37 %).

В качестве примера рассмотрим, какие статьи wikipedia.org привели пользователей на сайт «Научное наследие России» (табл. 9).

Таблица 9. Статьи сайта wikipedia.org, ссылающиеся на сайт «Научное наследие России»

№ п/п	Персоналии	%
1	Ковалевская, Софья Васильевна	40,02
2	Лобачевский, Николай Иванович	7,97
3	Жуковский, Николай Егорович	7,32
4	Чебышёв, Пафнутий Львович	10,47
5	Ляпунов, Александр Михайлович	5,93
6	Гмелин, Самуил Готлиб	4,36
7	Остроградский, Михаил Васильевич	3,99
8	Кёльрейтер, Йозеф Готлиб	1,76
9	Козлов, Пётр Кузьмич	1,76
10	Введенский, Борис Алексеевич	1,48
11	Урысон, Павел Самуилович	1,20
12	Ферсман, Александр Евгеньевич	1,20
13	Ориген	0,65
14	Грибовский, Вячеслав Михайлович	0,46
15	Фусс, Николай Иванович	0,37
16	Щербаков, Сергей Васильевич (1859–1932)	0,37

№ п/п	Персоналии	%
17	Гольмстен, Адольф Христианович	0,28
18	Кондратьев, Виктор Николаевич	0,28
19	Соколов, Михаил Андреевич	0,28
20	Шмидт, Фёдор Богданович	0,28
21	Гордин, Аба Лейбович	0,19
22	Арсеньев, Юрий Васильевич	0,09
23	Севергин, Василий Михайлович	0,09

Важно отметить, что среди учёных, приведённых в таблице, преобладают математики и ученые, специализировавшиеся в различных областях естественных наук. Это объясняется тем, что фонды электронной библиотеки «Научное наследие России» содержат оцифрованные версии редких научных изданий, которых нет на других ресурсах.

Поисковые запросы

Система Google Analytics выявила 17 338 вариантов поисковых запросов. Приводим двадцать наиболее употребляемых ключевых слов и словосочетаний, с помощью которых посетители искали информацию на нашем сайте (табл. 10).

Таблица 10. Наиболее употребляемые ключевые слова и словосочетания, используемые для поиска на сайте «Научное наследие России»

№ п/п	Запрос	Число запросов (% от всего числа запросов)
1	Плеханов	528 (1,90 %)
2	Аксельрод	253 (0,91 %)
3	Euleri	172 (0,62 %)
4	Соловьев	109 (0,39 %)
5	Логика	68 (0,24 %)
6	Правовое бытие	51 (0,18 %)
7	Папкович	46 (0,17 %)
8	Психология	85 (0,3 %)
9	Снегирев	43 (0,15 %)
10	Потебня	30 (0,11 %)
11	Власов	28 (0,10 %)
12	Лазер	28 (0,10 %)
13	Милович	28 (0,10 %)
14	Паллас	26 (0,09 %)

№ п/п	Запрос	Число запросов (% от всего числа запросов)
15	Фихтенгольц	51 (0.18 %)
16	Рынин Николай Алексеевич	25 (0.09 %)
17	Каптерев	24 (0.09 %)
18	Журнал прикладной химии	23 (0.08 %)
19	Ломоносов	23 (0.08 %)

Для поиска информации на нашем сайте посетители использовали фамилии авторов, названия изданий, названия областей науки и другие ключевые фразы. Особо выделяется третий по популярности поисковый запрос, который сделан в латинской транскрипции – Euleri (172 раза). Этот запрос принадлежит пользователям из дальнего зарубежья. На русском языке Эйлер был введен в строку поиска всего 6 раз. Подобный интерес обусловлен тем, что электронная библиотека «Научное наследие России» содержит самую большую в России (и возможно в мире) коллекцию работ Леонарда Эйлера.

Заключение

В заключение, можно попробовать представить портрет среднестатистического посетителя ТЭК нашего сайта. Как правило, это русскоговорящий житель крупного города СНГ в возрасте от 23 до 45 лет, он серьезно занимается историей науки и культуры (на нашем сайте нет развлекательных изданий). Впервые он обратился к сайту «Научное наследие России» в поисках редкого издания, которое отсутствует на других ресурсах (или возможно представлено не полностью в открытом доступе), и теперь регулярно обращается с запросами, используя сайт «Научное наследие России» как универсальный портал доступа к знаниям. При этом он все чаще и чаще применяет мобильные устройства для доступа к контенту нашей библиотеки.

Полученный среднестатистический (обезличенный) образ – лишь один из подходов к анализу работы библиотеки. Для продуктивного анализа работы ЭБ необходимо сегментировать входящие потоки запросов на наш сайт, выделять из общего потока отдельные группы наших посетителей по научным интересам, возрасту, полу, местонахождению и т.п. Чтобы получить расширенную статистику по запросам посетителей нам необходимо более тесно интегрировать модули Google Analytics и сайта «Научное наследие России». Для этого необходимо производить правильную установку целей анализа, добавляя на сайт блоки обратной связи. Все это подразумевает более тесное сотрудничество и взаимодействие между администраторами сайта и программистами.