

Библиопортал: что ожидает пользователь?

Достовалов Сергей Сергеевич, программист, Институт корпоративных библиотечно-информационных систем Санкт-Петербургского государственного политехнического университета

Постепенно традиционные походы в библиотеку сменяются “походами” на портал библиотеки. Причин этому несколько:

1. Оперативность. Зайти на портал гораздо быстрее, чем дойти до библиотеки.
2. Всё больше библиотек организуют электронные библиотеки - можно получить доступ к большому количеству материала “не выходя из дома” и не беспокоя сотрудников библиотеки.
3. Люди всё больше переносят свою жизнь в интернет. Интернет - эффективный канал обмена информацией, а информация - “конёк” библиотеки.

Зачем пользователь заходит на портал:

1. Получить первичную информацию о библиотеке.
2. Найти необходимый материал.
3. Воспользоваться сервисами: бронирования и заказа, чтение найденного материала, персонального информирования (“книги на руках”, консультации) и т.п.

Хорошо, когда пользователь находит нужную информацию. Плохо - когда пользователь нашел нужную информацию с трудом, а еще хуже, когда он ее не нашел вовсе (хотя она присутствует).

Как сделать так, чтобы пользователю было хорошо? Какие вопросы нужно задавать при проектировании и ведении портала? Как получить обратную связь от пользователя, информирующую о том, хорошо ли ему работается на портале?

Портал можно рассматривать с разных точек зрения, и каждая точка зрения порождает свой вопрос:

1. Портал, как структурированное хранилище информации. Вопрос: логична ли структура, удобно ли по ней перемещаться, нужна ли вообще эта информация?
2. Портал, как интерфейс пользователя. Вопрос: удобно ли пользоваться информацией, которая предоставляется через интерфейс портала?
3. Портал, как средство доступа к сервисам библиотеки. Вопрос: отвечает ли сервис портала поставленной цели, можно ли сделать использование сервиса более эффективным?

Каким образом можно получить ответы на поставленные вопросы? Ответ логичен - проследить за пользователем. Возникает вопрос: каким образом проследить за пользователем, какие инструменты выбрать?

Типы статистических инструментов

Статистические инструменты можно разделить на два типа:

1. системы общего назначения - фиксируют перемещение пользователя на портале;
2. системы специального назначения - анализируют действия пользователя с точки зрения бизнес-процессов, заложенных в функциональность портала.

Системы общего назначения

Наиболее популярные системы это “Яндекс.Метрика” и “Google Analytics”. Достоинства этих систем в том, что они предоставляют данные, которые возможно получить только в случае массового слежения за пользователями.

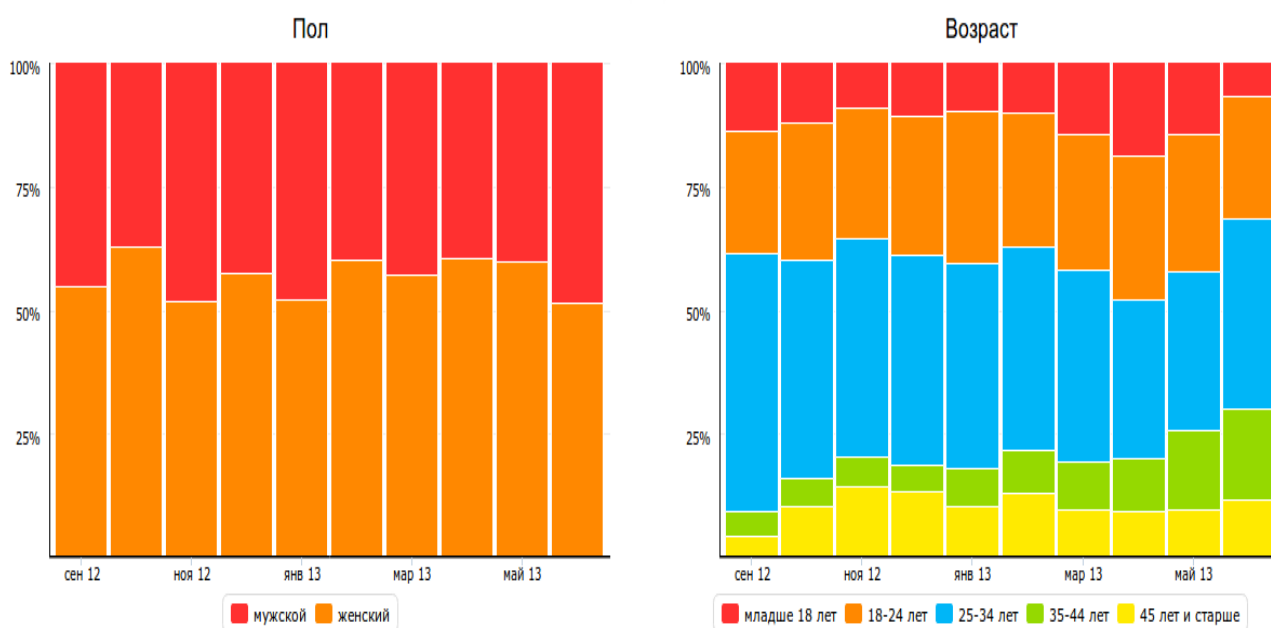


Рис.1. Половозрастная статистика посетителей

Категория	Аффинити-индекс
1. Литература и учебные материалы	163,9%
2. Недвижимость	130,2%
3. Семья и дети	126,3%
4. Обустройство	100,4%
5. Бизнес	96,0%
6. Программное обеспечение	89,3%
7. Кулинария	87,0%
8. Мобильная связь и интернет	83,1%
9. Финансы	82,3%

Рис. 2. Статистика по коммерческим интересам пользователей

Данные о пользователях, приведенные на рисунках 1 и 2, были получены на основании анализа посещения других сайтов.



Рис. 3. Тепловая карта кликов

На рис. 3. изображена тепловая карта кликов. Эта карта позволяет выявить самые востребованные (горячие) места портала. Возможно, некоторые элементы интерфейса нужно поменять местами?

В изучении поведения пользователей на сайте помогают новые инструменты: “Вебвизор” и “Аналитика форм” системы Яндекс.Метрика. Они позволяют записывать движения мыши пользователя и нажатие клавиш. Эти инструменты позволяют “подсматривать” за пользователем и анализировать его реакцию на события интерфейса портала.

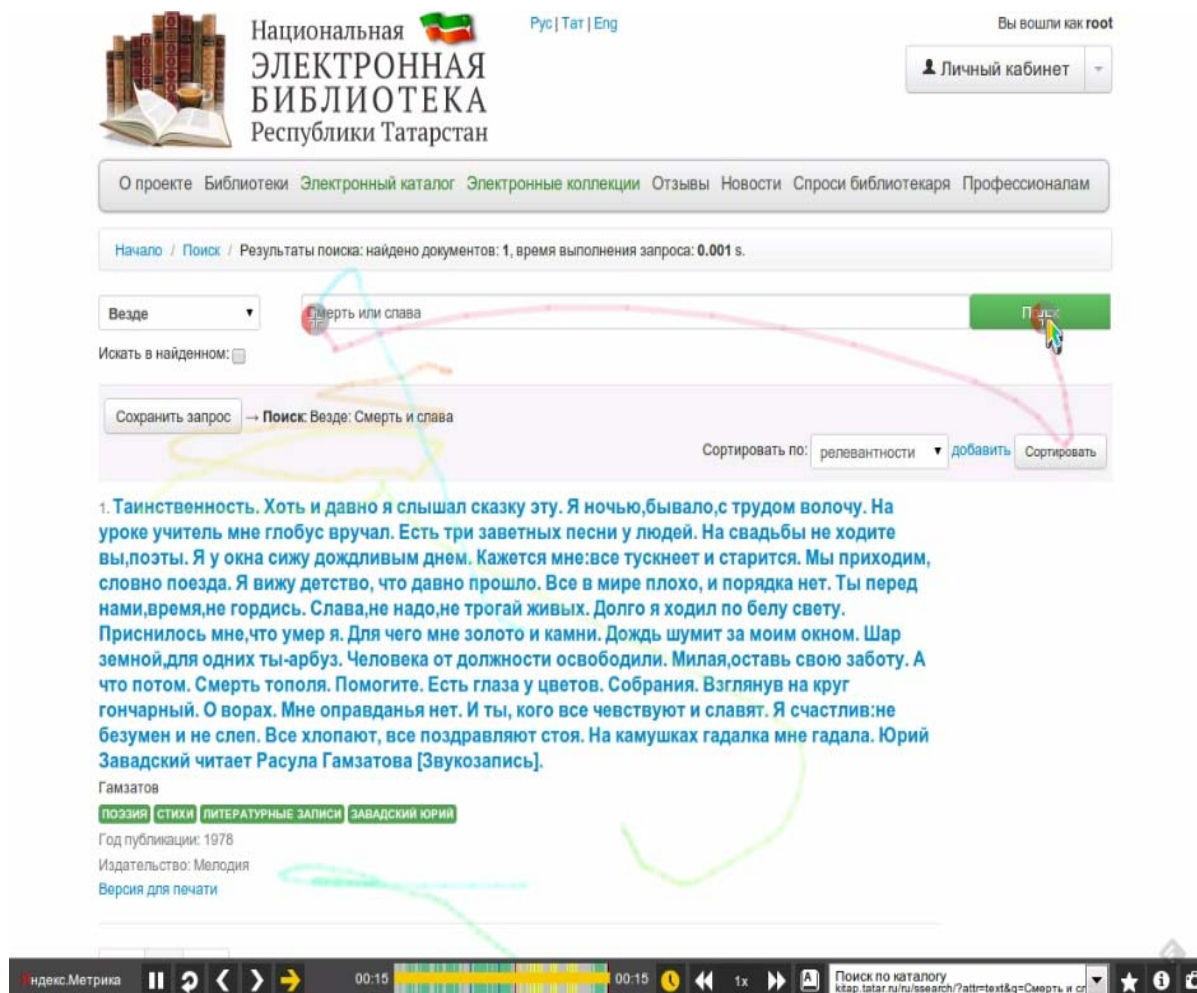


Рис. 3.1. Окно Вебвизора

На рисунке 3.1 показан интерфейс Вебвизора, на котором воспроизводятся все действия пользователя. Наблюдая за поведением пользователя можно сделать выводы о том, что ему было удобно, или наоборот - какие элементы интерфейса вызвали трудности или непонимание механизма их работы.

Системы специального назначения

Системы типа Яндекс Метрики или Google Analytics дают общую картину посещения портала пользователями, не вникая в особенности сервисов, предоставляемых порталом. И как бы они не были хороши (в этом нет сомнений), они не могут дать полной картины происходящего на портале, так как в них не заложены механизмы учета специфики бизнес логики сервисов.

Не для кого не секрет, что самым востребованным сервисом библиотечного портала, является поиск по каталогам и библиографическим базам данных. Получение и анализ статистических данных именно этого сервиса вызывает наибольший интерес, так как эти данные непосредственно влияют на качество и эффективность портала и позволяют повысить релевантность результатов

Цель анализа поискового сервиса:

1. Анализ покрытия каталога необходимыми поисковыми атрибутами: выявление атрибутов записи, которые хранят интересующую пользователя информацию.
2. Настройка поискового алгоритма в соответствии с интересами пользователей.
3. Получение сведений об интересах пользователей для повышения качества поисковой выдачи.

Анализ покрытия каталога необходимыми поисковыми атрибутами

Цель анализа:

1. Получить список наиболее популярных атрибутов, используемых пользователями.
2. Найти потенциально возможный атрибут, который был бы полезен для пользователей, для добавления в интерфейс или комбинирования уже существующих.
3. На основании проведенного анализа принять меры по улучшению качества поиска и подсчета коэффициента релевантности.

Получение списка наиболее популярных атрибутов

Приведенные диаграммы отображают качественную характеристику. Соотношения параметров не зависят от временного интервала.



Рис. 4. Распределение поисковых запросов по атрибутам (интерфейс вынуждает пользователя искать в определенном атрибуте)

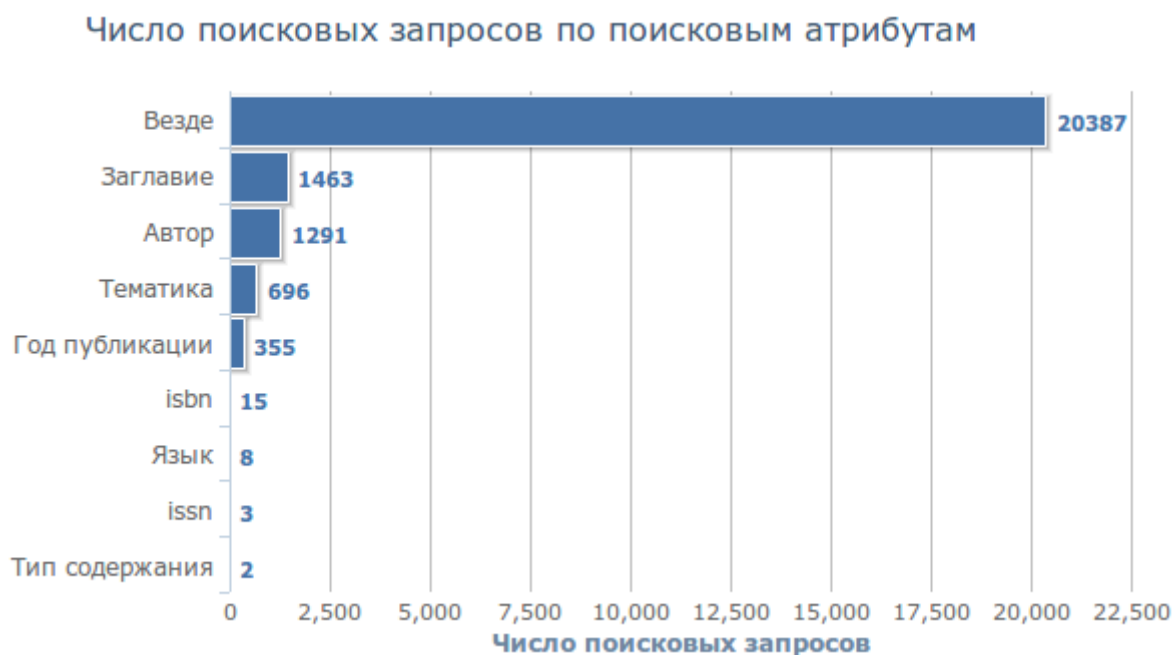


Рис. 5. Распределение поисковых запросов по атрибутам (интерфейс предоставляет пользователю единую поисковую строку с возможностью уточнения атрибута)

На рисунках 4 и 5 приведены диаграммы распределения поисковых запросов по атрибутам

На рисунке 4 статистика подсчитывалась на основании поискового интерфейса, “вынуждающего” пользователя указывать атрибут, по которому необходимо произвести поиск. Полученные значения соответствуют классическому распределению, где с большим отрывом лидирует 4 атрибута: автор, заглавие, тематика/ключевые слова, год публикации.

Диаграмма на рисунке 5 строилась на основании данных поискового интерфейса, предоставляющего единую поисковую строку (стоит заметить, что он так же позволяет выбирать поисковые атрибуты). Распределение поисковых атрибутов открывает совершенно иную картину использования поискового сервиса, которая является хорошим показателем того, как меняется стратегия поиска в зависимости от интерфейса. Общее количество запросов по атрибуту “Везде” (этот атрибут отображается единой поисковой строкой) составляет примерно 81%. Это говорит о том, что пользователи, в целом, удовлетворены результатами поиска, и им не требуется использование дополнительных атрибутов (соотношение запросов, которые вернули пустой результат одинаково в для первой и второй диаграммы).

Поиск потенциально возможных поисковых атрибутов

Диаграмма на рисунке 5 отлично демонстрирует самодостаточность атрибута “Везде”. С одной стороны хорошо, что пользователь может искать нужную ему информацию максимально быстро и просто, с другой стороны, такая простота накладывает дополнительные проблемы на поисковый сервис:

- Какую информацию учитывать при поиске по атрибуту “Везде”
- Включена ли вся необходимая информация, или наоборот, ее слишком много (что ведет к повышению процента поискового шума).
- Каким образом соотносить эту информацию при построении результата поиска.

В условиях, когда нет информации об атрибуте, в котором хотел бы искать пользователь (ему неудобно/невозможно об этом сообщать или он не знает), единственным выходом узнать об его желании является анализ его поисковых запросов непосредственно по фразам.



Рис. 6. Распределение поисковых запросов по фразам в атрибуте “Везде”

На рисунке 6 приведена статистика поисковых фраз. С уверенностью можно сказать, что наиболее часто встречающиеся поисковые фразы содержатся в заглавии, авторе, тематике.



Рис. 7. Распределение поисковых запросов по фразам в атрибуте “Автор”

Когда в распоряжении есть информация об атрибуте, можно с легкостью узнать, что интересует пользователей больше всего.

Меры, принимаемые для улучшения качества поиска

Каким образом можно использовать полученные данные:

- Динамическая смена приоритетов атрибутов. Поиск по атрибуту “Везде” подразумевает поиск одновременно по нескольким атрибутам. И если пользователи стали чаще делать запросы, которые содержат фамилию автора, то при подсчете коэффициента релевантности, приоритет отдается тем записям, которые содержат поисковую фразу в атрибуте “Автор”. В качестве примера можно привести поиск двух книг: “Биография Пушкина” и “Руслан и Людмила”. Если пользователь в поисковой строке введет фразу “Пушкин”, то с вероятностью, которую можно наблюдать в статистике атрибутов, пользователя интересуют записи, автор в которых указан “Пушкин”, следовательно, запись “Руслан и Людмила” попадет в поисковую выдачу с большим приоритетом, нежели “Биография Пушкина”.
- Составление профиля пользователя. На основе результатов поиска и просматриваемых документов делается картина предпочтений пользователя. Если пользователь интересуется информационными технологиями, то система в первую очередь выдает результаты, которые могут быть полезны для него. При построении картины анализируются тематические характеристики просматриваемых записей, и тело поискового запроса несущую смысловую нагрузку (учитываются только существительные). В случае, если пользователь связан с другими пользователями по какой либо характеристике (группа, курс, факультет и т.п.), то ему могут так же быть предложены записи, наиболее часто просматриваемые “похожими” пользователями.
- Оценить интересы пользователей. Подготовить более качественное описание ресурсов, которые наиболее часто ищут.

Заключение

Портал библиотеки неизбежно превращается из второстепенного в основной инструмент обслуживания пользователей. Классические отчеты, отображающие количество посещений пользователей, не достаточны для анализа использования портала, так как они отображают лишь малую часть данных, которые могут быть использованы в повышении качества обслуживания.

Описанные в докладе методы и инструменты статистики позволяют оценить поведение и информационную потребность пользователей. Анализ статистики не должен быть разовым, а проводиться с определенной периодичностью. Возможность использования статистических данных при поиске информации должна поднять работу с электронными каталогами на новый уровень, при котором информация в библиотечных базах данных адаптируется под пользова-

теля, а не пользователь под нее. Использование связей между пользователями для ранжирования результатов, помогают искать материал наиболее интересный целевой группе пользователей.