КРАЕВЕДЕНИЕ И ИНТЕРНЕТ

Сальников Сергей Георгиевич, *канд. физ.-мат. наук, вед. науч. сотр. Всероссийского института аграрных проблем и информатики им. А. А. Никонова*

Появление Интернета сделало проблему доступа к информации (в том числе краеведческой) и проще, и сложнее. Проще потому, что информации стало много больше, и она стала более доступной. Сложнее потому, что в легко обнаруживаемом огромном массиве информации по конкретной проблематике необходимо не только уметь ориентироваться, но и уметь давать самому себе ответы на другие вопросы: как найти релевантную информацию? Как её сохранить, преобразовать и эффективно использовать? Где нужные источники, и на каких условиях ими можно пользоваться? Попробуем ответить на эти и другие смежные вопросы достаточно подробно на примере сбора автором информации по конкретному населённому пункту, его малой Родине – станице Боргустанской (Предгорный район Ставропольского края).

С самого начала цель проекта было обычной, вполне приземлённой и, как казалось тогда, когда этот проект начинался (2009 г.), соответственно, и легко достижимой, а именно создание страницы в Википедии (https://ru.wikipedia.org/wiki/), посвящённой моей малой Родине. На момент начала работы над проектом страница эта фактически была лишь неким черновиком будущей возможной страницы и требовала очевидной переработки и дополнения.

Реализация проекта изначально предполагалась в онлайн-режиме. А именно был создан сайт, посвящённый проекту (http://borgustan.blogspot.ru/), и на этом сайте стала собираться информация по данному вопросу. Почему было принято такое решение? На то было несколько веских причин:

- бесплатность: не было (и нет) необходимости оплачивать услуги данного сервиса, с другой стороны, ограничений на количество размещаемого материала в настоящее время на этом ресурсе практически нет;

- высокая популярность: на данном ресурсе реализованы миллионы аналогичных (личных) проектов, которые прекрасно функционируют в последние годы;

- хорошая функциональность: помимо обычного наполнения и размещения текстовых страниц, размещаемый материал можно форматировать, снабжать метками (тегами), представлять в виде каталога (карта сайта), дополнять изображениями, видео- и аудиоматериалами и т.п.

- смысловое соответствие: публикация в виде отдельных сообщений (записей) вполне подходили для первоначального накопления материала, снабжение их тегами позволяло быстро (и достаточно наглядно) в полуавтоматическом режиме (теги приходится ставить самостоятельно, их организация в виде т.н. «облака тегов» - работа системы управления данным сайтом) строить их смысловую классификацию.

Для поиска источников данных первоначально использовался обычный поисковый запрос по ключевым словам, прежде всего по названию самой малой Родины. Источников данных набралось много, и можно было провести их некоторую предварительную систематизацию:

- официальные источники: это сайты библиотек, архивов, тематические сайты, специализированные проекты;

- неофициальные источники: сайты, наборы ссылок, форумы и т.п. ресурсы энтузиастов.

Официальные проекты (даже самые крупные) не гарантируют (потенциальную) полноту представленной информации. Так, например, на сайте Российской государственной библиотеки (http://www.rsl.ru/) можно найти (на конец ноября 2014 г.) информацию о том, что в её фондах имеется более 29 млн ед. хранения. В то же время на странице электронной библиотеки сайта можно найти информацию о 900 тыс. ед. хранения. Таким образом (нетрудно посчитать), даже ресурсы крупнейшей российской библиотеки оцифрованы только примерно на 3% (!) Что уж говорить о других проектах и ресурсах!

Есть и приятные исключения из правил. Так, например, для первоначального наполнения описываемого в этой статье проекта оказался полезным сервис Google Books (http://books.google.ru/books?hl=ru). Цель этого – создать его фактически полную электронную версию всех бумажных архивов человечества. Уже сейчас в электронных архивах этого проекта можно найти ссылки на тысячи книг на русском языке (в том числе и дореволюционных лет издания) в электронном формате (в основном это файлы самого популярного PDF-формата). Уникальность проекта в том, что многие книги не только оцифрованы, но и распознаны (!). Иными словами, по данным книгам организо-

ван исчерпывающий (т.е. полнотекстовый) поиск. Надо ли говорить, насколько это упрощает жизнь тем, кто собирает информацию по какой-либо определённой тематике?!

Другой пример – сайт Научной Электронной Библиотеки (http://elibrary.ru/). На нем можно найти более 6 млн научных публикаций российских авторов, полнотекстовые версии 4 тыс. иностранных и 4 тыс. российских журналов (20 тыс. выпусков). Доступ к 2800 российским журналам – открытый (!). Полноценный поиск организован по всей библиотеке электронных научных ресурсов. Здесь автору также удалось найти немало интересных материалов, хотя их научный характер, конечно, несколько ограничивал возможности их последующего использования для поставленных автором целей сбора такой информации.

Тематический сайт «Руниверс» (http://www.runivers.ru/) также может служить примером полноценного и очень полезного источника сведений исторического характера. Доступ к информации здесь также бесплатный (хотя есть определённые ограничения на использование информации, но они в целом не обременительны), полноценный поиск организован достаточно удобно, найденные источники можно скачать в нескольких наиболее популярных форматах для файлов текстового типа (PDF, DJVU и др.) или просмотреть их непосредственно на сайте в режиме онлайн. Среди недостатков проекта отметим его всё ещё сравнительно небольшой объём наполнения первоисточниками и отсутствие полнотекстового поиска по материалам этого сайта.

Сайты российских архивов в целом производят неудовлетворительное впечатление. Нужной краеведу информации на таких сетевых ресурсах (особенно на сайтах местных архивов) мало, электронные каталоги, как правило, отсутствуют. Единственная возможность найти что-либо нужное автору в таких архивах - направиться непосредственно в эти архивы, что, конечно, в большинстве случаев было невозможно географически и/или экономически. Неофициальные источники информации оставляют противоречивые впечатления. С одной стороны, только на этих ресурсах можно найти многие материалы, которых на официальных сайтах просто нет. С другой – понять, насколько правомочно использование таких источников практически невозможно. Всё же представляется, что информацией из дореволюционных источников можно в большинстве случаев пользоваться по причине истекшего срока давности. Использование послереволюционных (тем более относительно современных) источников, думается, допустимо только на уровне ссылок или небольших цитат из текста с указанием источника цитаты.

Другая беда неофициальных источников: попытки коммерционализации и вредоносная деятельность (распространение вирусов). Здесь посетителям таких сайтов можно посоветовать только одно: не платить, если не уверен; не качать, если не уверен; иметь всегда в наличии (в рабочем, активном состоянии) на своём компьютере во время работы с такими сайтами хороший антивирусный пакет. Крайне не рекомендуется (во избежание серьёзных финансовых потерь) предоставление на таких сайтах какой-либо информации о себе (особенно информации о номерах мобильных телефонов). Иначе проблем с источниками такой информации не избежать.

Найти нужные источники информации по интересующей проблеме обычно не составляет особого труда, несколько запросов по ключевым словам обеспечат исследователя первоначальной информацией. Наиболее популярные здесь средства хорошо известны – это поисковые службы (прежде всего - Google и Яндекс). Наиболее развитый сервис такого рода – это, конечно, первый из названных, но отечественным сервисом Яндекс также не следует пренебрегать. Уже одно то, что он входит в мировую пятёрку самых популярных в мире поисковиков, говорит само за себя. Наличие тонких (языковых) настроек при поиске информации – тоже немаловажное конкурентное преимущество Яндекса. Таким образом, искать информацию можно (и нужно) в Google, но затем её необходимо дополнить (а может быть, и уточнить) при помощи Яндекса.

Самая большая проблема – наличие в результатах выдачи любого поискового сервиса по любому ключевому слову т.н. «нерелевантной информации» (т.е. информации, к делу не относящейся и подходящей запросу только формально). На самом деле устраняется при использовании специальных средств поиска. Под этим понимается не только т.н. «расширенный поиск» (который позволяет снизить количество нерелевантных ссылок на порядок), но и специализированные средства поиска типов: «поиск точного совпадения», «поиск словосочетания», «поиск всех совпадений во фразе», «поиск точной цитаты» и т.п. Все эти средства т.н. «точного» поиска в каждом поисковом сервисе реализуются по-своему. Справочная информация в каждом хорошем поисковике обширна и даёт исчерпывающий ответ (в соответствующих разделах) на

то, как организовать специализированный поиск. Изучите её, и тогда процедура поиска нужной информации станет не только не утомительной, но и приятной.

Следующая поисковая проблема: что можно найти в Интернете (по типу информации и/или электронному формату)? Опыт автора данной статьи показывает, что при поиске краеведческой информации речь может идти только о нахождении текстовой информации. Видео и звуковые файлы на эту тему в Интернете почти не представлены, изображения, как правило, грешат плохим качеством. По этой причине, например, карты, найденные в Интернете пригодны только для предварительного ознакомления с нужной тематикой.

Текстовая краеведческая информация (книги, брошюры, статьи и т.д.) по любой тематике представлена в Интернете достаточно широко. Базовый формат здесь, конечно, формат типовых страничек Интернета (стандартный для страниц сайтов HTM- и HTML-формат представления данных). Хотя в последнее время отчётливо намечается тенденция замещения такой информации на простые ссылки на текстовые файлы других форматов (как правило, это самые популярные форматы такого типа – PDF или DJVU) или на ссылки на страницы сайтов, где такие книги можно прочитать в онлайн-режиме. Причина таких изменений не только в том, что предлагаемая организация материала гораздо проще реализуется. Немаловажно, что появились и типовые средства просмотра в онлайн-режиме таких файлов, что снимает проблему создания красивых (но очень затратных по времени и стоимости их создания) html-страниц с текстом, «вынутым» из соответствующей книги (монографии).

Не следует забывать о том, что (в отличие от html-страниц) текстовую информацию из файлов PDF- или DJVU-формата извлечь (напрямую и «задёшево») можно далеко не всегда. Если PDF-файл формировался из страниц, которые были только оцифрованы, но не распознаны, то такой файл, т.н. «текстового слоя» содержать не будет, и поиск необходимой информации в нём будет возможен только чисто визуально.

Конечно, можно самостоятельно организовать распознавание нужного материала, но это отдельная проблема.

Следующая проблема, возникающая после сбора всего необходимого для изучения электронного материала на одном (локальном) компьютере, такова: как эффективно работать с этим (предположительно) объёмным массивом информации? Конечно, систематизация по году издания, авторам, названиям, ключевым словам и т.п. – дело обычное, и нет смысла на этом останавливаться в данной статье. Существуют ли (полу)автоматические средства обработки массивов текстовой информации, которые бы превращали их в полноценные архивы данных? Имеется в виду, прежде всего возможность полнотекстового поиска по всему такому массиву по определённым (ключевым) словам или их сочетаниям, т.е., образно выражаясь, можно ли создать на своём компьютере небольшой/личный/локальный Яндекс? Оказывается, такие средства уже давно существуют и вполне доступны. Так, автор этой статьи много лет пользуется одной из лучших (по соотношению цена/качество) систем такого класса программного обеспечения, а именно системой “Архивариус 3000”. Невысокая текущая стоимость этой системы (http://www.likasoft.com/ru/order.shtml) – 395 руб. для персональной редакции – открывает возможность эксплуатации программного продукта, обладающего рядом привлекательных характеристик. А именно: полнотекстовый поиск в документах и электронной почте, поддержка различных кодировок, морфологический поиск, индексация более 40 самых распространённых видов документов, индексация архивных файлов (без их разархивации!), наличие серверной версии и удалённого доступа и т.д.

Личный архив документов благодаря пакетам типа «Архивариус» перестаёт быть вещью в себе: по любому слову (или фразе) можно мгновенно получить список всех документов вашего личного электронного архива, в которых встречается это слово (или фраза). Здесь же можно получить и экранное изображение той части текста, где встретился поисковый запрос. Это позволяет быстро оценить возможности личного архива для сбора информации по определённой теме, создать необходимую для этого подборку материалов из личного архива и приступить к работе с данным массивом.

В каком состоянии в настоящее время находится проект по моей малой Родине – станице Боргустанской (http://borgustan.blogspot.ru/)? На сайте около 900 сообщений по самым разным тематикам – история и экономика, география и геология, этнография и статистика и т.д. Все сообщения (статьи блога) помечены релевантными метками (тегами). По облаку тегов на главной странице сайта можно легко найти нужные мате-

риалы. Имеется также карта сайта, где страницы сгруппированы по упомянутым выше тегам. Наконец, на сайте работает типовой поиск (от владельца сервиса – фирмы Google) по всем сообщениям проекта. Так что (в случае необходимости) поиск информации на сайте можно осуществить и типовыми средствами поиска Google.

Кроме того, у автора собран архив из нескольких тысяч единиц электронных материалов, организованных при помощи программного пакета «Архивариус» в полноценный локальный архив. Сайт проекта (http://borgustan.blogspot.ru/) регулярно пополняется вновь найденными материалами из этого архива. По результатам поиска Google по ключевым словам, относящимся к данной тематике, сайт проекта неизменно оказывается на первой странице выдачи поискового сервиса Google.

Что интересного удалось собрать по истории жизни станицы Боргустанской?

1. Найдены копии подлинных документов-приказов по истории основания станицы знаменитым генералом А.П. Ермоловым (http://borgustan.blogspot.ru/2011/07/3.html):

2. Материалы по истории слова «Боргустан», связанного с существованием на месте первоначального основания станицы древнего города-крепости «Боргустан» (Бургустан, Бергусант). В том числе и упоминания таких известных путешественников по этим местам, как П.-С. Паласс, Я. Потоцкий, Эвлия Челеди и др. Фильм об этом древнем историческом памятнике недавно появился в сети в открытом доступе (см., например, http://borgustan.blogspot.ru/2013/02/62.html);

3. Многочисленные упоминания станицы в классическом труде Потто по Кавказской войне (http://borgustan.blogspot.ru/2011/02/blog-post_08.html);

4. Многочисленные статистические сведения по станице в различных сборниках и календарях (Кавказский календарь, Терский календарь и т.д. см. например, http://borgustan.blogspot.ru/2014/07/1892.html);

5. Материалы по Гражданской войне (мемуары Шкуро http://borgustan.blogspot.ru/2009/12/10.html, Врангеля, Миронова и др.);

6. Сведения о знаменитых земляках, их которых самым известным был, конечно, Д.Р. Ветренко (http://borgustan.blogspot.ru/2010/01/blog-post_4901.html), дивизия которого во время осеннего 1919 г. наступления Юденича на Петроград получила приказ перерезать Николаевскую железную дорогу в районе станции Тосно. По неясным причинам приказа не был выполнен, что позволило красному командованию быстро перебросить к Петрограду многочисленные резервы из Центральной России. Невыполнение Д.Р. Ветренко этого приказа считается одной из основных причин провала наступления белых на Петроград;

7. Собран и опубликован полный свод сведений с сайта «Мемориал» по станичникам – жертвам репрессий 20-40 гг. XX века (http://borgustan.blogspot.ru/2010/02/blog-post_7969.html);

8. Выложен на сайт главный краеведческий труд местного масштаба – работа о станице бывшего её жителя Г.И. Зинченко (http://borgustan.blogspot.ru/2011/07/blog-post.html);

9. Опубликовано много материалов из Google Books, их можно скорее считать основой для дальнейших изысканий, чем источником полной информации, но всё же ценной и обширной по тематике и охвату;

10. Сведения по географии, геологии, экономике, этнографии и т.д., относящиеся к станице.

В ближайших планах – систематизация материалов, создание сводных тематических страниц на сайте (в том числе и в первую очередь страница хронологии жизни станицы за все годы её существования) и …постепенное наполнение страницы Википедии, посвящённой станице.

МЕЛЬНИЦА МОЕГО ДЕТСТВА

***Потапов Александр Николаевич**, член Союза писателей и Союза журналистов России (г. Рязань)*

Каждый, кто проезжает по трассе Москва – Самара, непременно обращает внимание на старую ветряную мельницу, которая стоит у шумной дороги в шацком селе Польное Конобеево, на 368-м километре от столицы. Говорят, этот ветряк – единственный не только в Рязанской области, но и в России.

Помнится, в годы студенческой юности я частенько наведывался из Рязани на родину, в село Польное Конобеево Шацкого района, где жили мои родители. Доезжал на городском автобусе до окружной дороги и шел на автозаправку. Как правило, там было немало машин, следующих по трассе Москва – Куйбышев из дальних городов и