

Открытые API и создание сети разработчиков электронных библиотек

В последние годы наблюдается постоянный рост количества разнообразных электронных коллекций, опубликованных в Интернете газетами, журналами, издателями, библиотеками, музеями, университетами и многими другими учреждениями науки и культуры. При этом, конечно же, неизбежно происходит пересечение интересов, одни и те же документы становятся интересны и уместны в различных коллекциях. Для реализации одного из основных принципов автоматизации – «однократного ввода и многократного и многоцелевого использования» – применяется технология web-сервисов, с помощью которой организация, создавшая ресурс, может предоставить возможность всем желающим интегрировать его в собственные ресурсы.

К сожалению, в России в сфере создания электронных библиотек эта тенденция выражена очень слабо. В настоящей статье мы попробуем проанализировать, как развивается эта деятельность в электронных библиотеках других стран, и что нужно сделать для того, чтобы и в России взаимосвязи электронных библиотек стали нормой, а не исключением.

Основной проблемой для создания web-сервисов всегда является неясность потребности в них. С одной стороны, зачем создавать сервис и документировать API¹ к нему, если ещё непонятно, найдутся ли разработчики, которые будут создавать свои приложения на его основе. Но, с другой стороны, никто и не будет создавать приложения, если ещё нет сервиса и API к нему. Поэтому создание сообщества разработчиков, развитие интереса к создаваемым сервисам и формирование спроса на них – одна из основных задач во всех успешных ресурсах современности.

Многие знают про работу, проводимую Google для привлечения разработчиков (<https://developers.google.com/>), но про аналогичную деятельность в сфере электронных библиотек известно меньше. Лидирующую позицию в этой области занимает OCLC². Часто задают вопрос, что является в этой ситуации первичным: создаётся сообщество, потому что корпорация успешна, или корпорация успешна, потому что создано сообщество разработ-

чиков? На мой взгляд, всё это тесно взаимосвязано. В последние несколько лет OCLC оказалась на новой ступени развития. Она перешла на платформу OCLC WorldShare, предоставив пользователям ряд web-сервисов и, что, может быть, самое интересное, открытых API и их подробных описаний для всех желающих на специально созданном для этого сайте <http://www.oclc.org/developer/>. Были представлены API к следующим сервисам:

- OpenURL Gateway – сервис переадресации к полнотекстовым ресурсам;
- QuestionPoint Knowledge Base – поиск в базе данных выполненных справок;
- WorldCat Search – поиск во всех записях WorldCat;

1. API (Application Programming Interface) – интерфейс программирования приложений (иногда – интерфейс прикладного программирования) – набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) для использования во внешних программных продуктах.

2. OCLC (Online Computer Library Center, INC) – организация, занимающаяся каталогизацией ресурсов для всего мира на разных языках, обеспечивающая библиотеки возможностью выгружать из её баз данных готовые библиографические данные и ставить их в свои каталоги (прим. ред.)

Автор



Михаил Ефремович ШВАРЦМАН

начальник отдела исследования компьютерных систем Российской государственной библиотеки, shvar@rsl.ru

- WorldCat Identities — поиск сведений о персоналиях (авторы, персонажи, корпорации и т.д.) в WorldCat;
- WorldCat Registry — поиск во всемирном справочнике библиотек, архивов, музеев;
- WMS Vendor Information Center — поиск информации о продавцах;
- xISBN (xISSN) — поиск по ISBN (ISSN) в WorldCat и ряду других.

Большое внимание уделяется в OCLC продвижению новых возможностей. Так, например, выделены сотрудники, ежедневно отвечающие за поддержку этих сервисов, и определены часы их работы по поддержке разработчиков. Регулярно проводятся семинары и вебинары. На международных конференциях, связанных с разработкой библиотечного программного обеспечения, при поддержке OCLC выделяются секции для обсуждения разработок приложений для работы с API OCLC.

Посмотрим, каков же результат этой работы. В настоящее время на сайте разработчиков опубликованы описания 110 приложений для различных сервисов. Для того чтобы все желающие могли убедиться, что ничего сложного в создании приложений нет, первые приложения были сделаны сотрудниками OCLC, а их исходные коды были выложены в открытый доступ. Среди таких приложений наиболее интересным мне представляется xISSN History Visualization Tool (<http://worldcat.org/xissn/titlehistory>), который использует сервис поиска метаданных по ISSN и строит схему слияний, разделений и переименований журнала. Пример работы этой программы для ISSN 0148-0227 приведён на рис. 1.

Больше всего приложений (45) создано для использования WorldCat Search API. Среди них присутствуют приложения для различных платформ. Так, например, сделано приложение для Facebook (<http://apps.facebook.com/worldcat/>), которое позволяет проводить поиск в базах данных OCLC, не выходя из Facebook (рис. 2), и обещает учитывать интересы пользователя, занесённые в его профиль. Справедливости ради следует отметить, что мне не удалось понять, как они учитываются, а кириллица при поиске работает некорректно. Но дело не в этом, а в том, что нашлись разработчики, осознавшие необходимость использования библиографического поиска и формирования библиографического списка из найденной литературы в соответствии с различными стандартами (русского, конечно же, там нет).

Очень важной для пользователей оказывается задача определения ближайшего местоположения книги. Для поиска изданий в российских библиотеках эту информацию можно получить на портале госуслуг <http://www.gosuslugi.ru> или в сводном каталоге «ЛИБНЕТ» <http://www.nilc.ru>, однако результаты поиска можно посмотреть только на этих сайтах. Никаких открытых средств интеграции поиска во внешние системы не предусмотрено. Видимо, поэтому никто и не разрабатывает приложений, учитывающих эти данные. В OCLC пошли другим путём. Все их сервисы открыты для некоммерче-

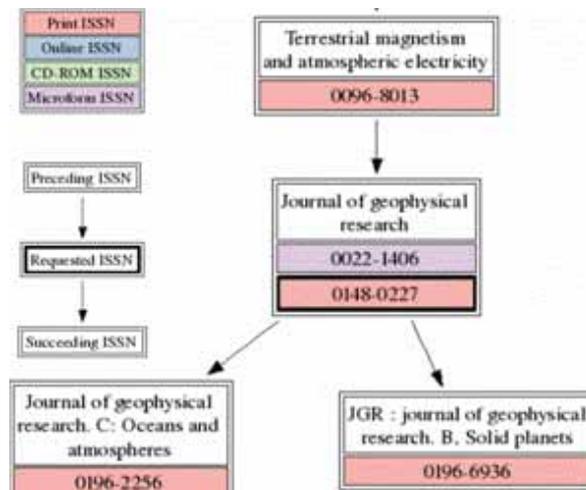


Рис. 1. Результат работы программы xISSN History Visualization Tool



Рис. 2. Приложение Facebook для поиска в OCLC



Рис. 3. Приложения для мобильных устройств



Balance sheet (London) (0965-7967) Peer Reviewed ✓
 from 01/01/1994 to 12/01/2004 in [ABI/INFORM Global](#)

BizEd (St. Louis, Mo.) (1537-338X)
 from 11/01/2001 to present in [Academic Search Complete](#), [Education Research Complete](#) and [Professional Development Collection](#)

Decision sciences journal of innovative education (1540-4595) Peer Reviewed ✓
 from 03/01/2003 to 1 year ago in [Business Source Complete](#) and [Education Research Complete](#)

Рис. 4. Приложение Peer Reviewer

The screenshot shows the Peer Reviewer application interface. At the top, there are search filters and a search button. Below, there are several search results for Yuri Gagarin, including a registration number (2004/71), production date (1962-1968), and acquisition credit line. A red circle highlights a 'Peer-reviewed' status indicator. Below the search results, there is an 'Overview' section for Yuri Alekseyevich Gagarin (1934-1968), showing 292 works in 402 publications in 31 languages and 4,453 library holdings. A 'Publication Timeline' chart shows the distribution of publications over time. A 'Most widely held works' section lists several books, including 'The first man in space' by Michael D. Cole and 'Stardust: the truth behind the legend of Yuri Gagarin' by Jamie Doran. An 'Alternative Names' section lists various names and dates associated with Gagarin.

Рис. 5 и 6. Сервис WorldCat Identities

The screenshot shows the Collins Catalog website interface. At the top, there is a search bar and navigation links. Below, there are search results for Yuri Gagarin, including a 'NY Times Hardcover Nonfiction bestsellers' section. A red circle highlights a 'NY Times Hardcover Nonfiction bestsellers' status indicator. Below the search results, there is a 'Collins Catalog' section with a list of books, including 'The first man in space' by Michael D. Cole and 'Stardust: the truth behind the legend of Yuri Gagarin' by Jamie Doran. A red circle highlights a 'NY Times Hardcover Nonfiction bestsellers' status indicator.

Рис. 7. Сервис библиотеки University of Puget Sound

ского использования всеми желающими. Поэтому приложения, решающие задачу определения местоположения книги, разрабатываются в большом количестве, различными программистами и с различным функционалом. Для решения такой задачи нужно

воспользоваться уже двумя сервисами: WorldCat Registry и WorldCat Search. Различных приложений, решающих эту задачу, сделано немало. Мне более других понравилось приложение Pic2Shop. Это бесплатное приложение для Android или iPhone, которое превращает ваш телефон в сканер штрихкода. Оно позволяет сравнить цены и проверить, есть ли в местной библиотеке книги, которые вас заинтересовали (рис. 3).

Для многих учёных очень важно знать, является ли тот или иной журнал рецензируемым. Потому задача использования сервиса xISSN привлекала нескольких разработчиков. В результате появилось несколько приложений, решающих одну задачу — простановку отметки о рецензируемости в списке журналов. Например, сотрудники IDS Project разработали приложение Peer Reviewer (<http://www.idsproject.org/Tools/PeerReviewer.aspx>), которое добавляет такие отметки (рис. 4).

Библиографические данные нужны не только пользователям библиотек. При описании музейных объектов бывает необходимо делать ссылки на книги, связанные с объектами, или на личностей, имеющих отношение к этим объектам. Для того чтобы использовать огромный запас данных, уже имеющийся в OCLC, музей Powerhouse (Австралия, <http://www.powerhousemuseum.com>) при описании личностей использует сервис WorldCat Identities. На рис. 5 мы видим, что при выборе фотографии Юрия Гагарина нам выдаётся ряд сведений о фотографии как объекте хранения, а за сведениями о Юрии Гагарине нас отсылают в OCLC.

Если мы проследуем по этой ссылке, то получим довольно подробную информацию о нём (рис. 6).

Хочу отметить, что этот музей не только сам использует сторонние web-сервисы, но и предоставляет свои (<http://api.powerhousemuseum.com>).

В некоторых библиотеках одновременно используются сервисы сразу нескольких ресурсов. Так, например, New York Times предоставляет API к своему набору сервисов <http://developer.nytimes.com/>. На основе этого API различными авторами был создан ряд приложений, улучшающих возможности использования ресурсов New York Times. Среди них есть возможность получения списка бестселлеров на текущий момент в различных по типу обложки или по жанру категориях книг.

В библиотеке University of Puget Sound объединили возможности этого API и сервисов OCLC. На сайте библиотеки можно выбрать категорию книг, в которой ищется бестселлер, в полученном от NY Times списке выбрать интересующий вас бестселлер и затем, используя WorldCat Search, получить название библиотеки, в которой имеется это издание (рис. 7).

Интересные приложения можно обсуждать долго, однако нас интересует вопрос, как мотивировать сторонних разработчиков в участии и создании различных приложений. Из опыта OCLC мы видим, что успех им принесли: серьёзное внимание со стороны сотрудников, хорошо документированные средства, постоянные мероприятия по обучению и реклама этой деятельности.

Однако OCLC — это некоммерческая организация; интересно оценить, как обстоят дела у коммерчески успешных компаний в нашей сфере.

Если мы посмотрим, как работает Elsevier, создавая своё сообщество разработчиков дополнительных приложений (<http://www.applications.sciverse.com/action/gallery>), то увидим, что в его подходе есть принципиальные отличия от OCLC. Приложения, в основном, нацелены на улучшение поисковых возможностей платформы Sciverse, на более удобную визуализацию полученных результатов, на получение дополнительной информации. Все приложения работают на платформе Sciverse, внутри браузера, а внешним системам не предоставляется доступ к данным. Для мотивации разработчиков используется, в основном, конкурс с хорошими денежными призами. Так, в последнем конкурсе (<http://appsforscience.com/>) было более 6000 участников, а призовая сумма составила 35 000 долларов (рис. 8). Различные мероприятия проводятся существенно реже, чем в OCLC, но примерно один раз в месяц в мире происходит какое-то мероприятие, посвящённое разработке приложений. Разумеется, у разработчиков есть свой сайт, на котором ведётся обсуждение возникающих проблем.

У компании EBSCO иной подход. Они разработали сервис EBSCOhost Integration Toolkit (<http://support.epnet.com/eit/>), в который входит в том числе и web-сервис, сделали к нему подробную документацию, и пока никаких действий по привлечению сторонних разработчиков не предпринимают.

Если мы проанализируем проекты разработчиков программного обеспечения для электронных библиотек, то некоторые из них действуют аналогично Elsevier. Например, компания Exlibris также организовала сообщество разработчиков (<http://www.exlibrisgroup.org/display/CodeShare/Home>), которое нацелено, в основном, на улучшение функционала продуктов компании. Но поскольку API к ряду своих продуктов компания опубликовала, разрабатываются и различные интеграционные приложения. Например,



Рис. 8. Конкурс разработчиков Apps for Science

Combine Primo Central API Records with Solr Records ([http://www.exlibrisgroup.org/display/PrimoCC/Combine + Primo + Central + API + and + Solr + Records](http://www.exlibrisgroup.org/display/PrimoCC/Combine+Primo+Central+API+and+Solr+Records)). Для мотивации разработчиков используются соревнования, так же как и в Elsevier. Правда, проводятся они реже, участников бывает меньше, а призовой фонд существенно скромнее — 12 000 долларов в 2010 г.

Оценивая состояние дел в России, можно сказать, что российские поставщики информации практически не предоставляют API к своим ресурсам. Даже те из них, кто даёт свободный доступ к метаданным через web-интерфейс, не предоставляют открытого API. Хотя, на мой взгляд, потребность у российских пользователей есть, и наверняка бы нашлись желающие воспользоваться такими API и написать свои приложения для интеграции внешних ресурсов и сервисов в свою информационную инфраструктуру.

В отсутствие такой возможности приходится использовать довольно трудоёмкий способ разбора HTML-страниц в поисковом интерфейсе нужного ресурса. Примером такого подхода может быть созданный в Сибирском федеральном университете (СФУ) механизм федеративного поиска. На странице <http://libsearch.sfu-kras.ru/> размещены форма поиска и список ресурсов, к которым университет имеет доступ. Для каждого ресурса заранее был изучен web-интерфейс его поисковой страницы и написана программа, автоматически заполняющая необходимые поля в поисковой форме и разбирающая ответ. Из российских ресурсов запросы от СФУ отправляются в следующие ресурсы:

- электронная библиотека Издательского дома «Гребенников»;
- электронная библиотека диссертаций РГБ;
- электронно-библиотечная система издательства «Лань».



МЕТАПОИСКОВАЯ СИСТЕМА ПО ЭЛЕКТРОННЫМ БИБЛИОТЕКАМ

*А.В. Спирина, инженер отдела АБИС
Ресурсного центра Библиотечно-издательского комплекса
Сибирского федерального университета, г. Красноярск*

Научно-образовательная сфера деятельности тесно связана с поиском информации в глобальной сети Интернет. Количество научно-образовательных информационных ресурсов стремительно растёт. Для поиска информации, как правило, каждый из таких ресурсов предлагает свою систему. За этим многообразием скрываются определённые проблемы поиска информации, связанные с тем, что пользователь вынужден совершать один и тот же запрос в каждом из ресурсов, затрачивая на это много времени. Кроме этого, они отличаются поисковыми интерфейсами, навигацией, правилами запросов. Во всём этом многообразии бывает достаточно трудно разобраться, что снижает эффективность отбора необходимых документов.

Научная библиотека Сибирского федерального университета (НБ СФУ) является одной из крупнейших вузовских библиотек в России. Она имеет доступ к онлайн-базам данных крупнейших российских и зарубежных производителей информационных продуктов, содержащим электронные полнотекстовые версии научных журналов, газет и книг, материалов конференций, патентную, библиографическую и наукометрическую информацию. Помимо внешних электронных библиотек, СФУ обладает собственными ресурсами: «Видео СФУ», «Архив электронных ресурсов СФУ», двумя электронными каталогами НБ СФУ.

Проблема поиска информации в многочисленных электронных библиотеках характерна не только для НБ СФУ, но и для других вузов и научных учреждений, имеющих доступ к различным электронным библиотекам.

В связи с этим было решено разработать метапоисковую систему по онлайн-ресурсам, которая бы позволила пользователю одновременно вести поиск во многих ресурсах через единый web-интерфейс и получать результаты поиска, сгруппированные по ресурсам, на одной web-странице.

Существуют специализированные интернет-протоколы для обмена информацией электронными библиотеками (Z39.50, OAI-PMH, SRU/SRW), однако далеко не все ресурсы их поддерживают. Несмотря на то, что специализированные протоколы являются более эффективными механизмами поиска, для разработанной метапоисковой системы был выбран протокол HTTP. Поскольку он является слишком общим, в нём нет стандартизированных типов запроса и каждый интернет-ресурс, работающий по данному протоколу, имеет свой формат запроса и ответа, вследствие чего в разрабатываемой поисковой системе необходима настройка профиля для каждой из электронных библиотек.

Кроме того, в случае изменения web-интерфейса какого-либо интернет-ресурса появится необходимость корректировки настроек профиля для данного интернет-ресурса.

Разработанная метапоисковая система по электронным библиотекам строится на принципах клиент-агент-серверной архитектуры, где в качестве клиента выступает web-браузер, в качестве агента – поисковая машина метапоисковой системы, а в качестве сервера – web-серверы электронных библиотек, к чьим поисковым механизмам обращается поисковая машина (агент).

Разработанная поисковая система находится по адресу <http://libsearch.sfu-kras.ru/> и работает следующим образом.

- Пользователь заполняет поисковую форму и отправляет запрос на web-сервер.
- Этот запрос (адаптированный / трансформированный под конкретную поисковую систему) с web-сервера отправляется в поисковые системы электронных библиотек (по протоколу HTTP).
- На стороне электронных библиотек при помощи поисковых механизмов самих этих библиотек происходит обработка запроса.
- Метапоисковая система получает HTML-коды web-страниц с результатами поиска, в которых содержатся метаописания найденных документов (автор, заглавие, аннотация, год издания, название журнала, где была опубликована статья и т.д.).
- На web-сервере происходит разбор HTML-кодов полученных страниц с помощью регулярных выражений для выделения нужной информации.
- Обработанная информация предоставляется по мере поступления результатов пользователю с разбивкой по электронным библиотекам.

Разработанная система позволяет осуществлять поиск по автору, заглавию, ключевым словам. Также существует возможность использования расширенного поиска, который добавляет к основному поиску поля типа «Год издания ___ по ___».

Данная поисковая система не поддерживает использование булевых операторов («И», «ИЛИ»). При отправке запроса в поисковые формы электронных библиотек, поддерживающие эту опцию (булевые операторы), в поисковых формах этих электронных библиотек автоматически выбирается оператор «И» (настройка производится для каждой конкретной электронной библиотеки при написании шаблона url-запроса и её можно изменить).

Поисковая форма разработанной системы предоставляет пользователю возможность выбора отрасли науки, после выбора отрасли система предоставляет пользователю список электронных библиотек, специализирующихся на этой отрасли.

В РГБ запросы можно отправлять по протоколу Z39.50, а для остальных двух ресурсов такой возможности нет. Конечно, было бы удобнее иметь стандартный протокол обмена данными со всеми издателями. В этом случае не нужно было бы беспокоиться о возможных изменениях в интерфейсе у издателя и при каждой смене интерфейса исправлять программу отправки запросов и разбора ответов. Я не идеализирую такой подход в организации метапоиска и понимаю все его недостатки, однако считаю, что во многих случаях он вполне работоспособен. Мне кажется, что это направление деятельности довольно интересное, опыт СФУ вполне может быть использован и другими университетами. Набор ресурсов, выписываемых российскими организациями, довольно типичен, и программы-разборщики, созданные для одного подписчика, вполне могут быть использованы и другими организациями.

Что касается российских разработчиков программного обеспечения для электронных библиотек, то здесь тоже не видно сообществ разработчиков вокруг компаний-производителей. Мне удалось найти только один пример такого сообщества. Это форум разработчиков для «ИРБИС» <http://irbis.gpntb.ru/list.php?24>. **М.В. Гончаров**, директор Центра развития и интернет-технологий ГПНТБ России, рассказывает: *«Кроме открытого форума, в котором обсуждается решение текущих трудностей, существует его закрытая часть, в которой участвуют только разработчики. Среди них большую часть составляют сторонние участники проекта из различных регионов России, Белоруссии, Украины и Казахстана, которые занимаются созданием дополнительных конвертеров, шаблонов для «ИРБИС» и т.п. программ, не связанных с перелкой ядра программы. В настоящее время более 20 % программ создано именно такими разработчиками.*

Поскольку для «ИРБИС» существует открытый API, его пользователями был разработан целый ряд программ, расширяющих или дополняющих основные модули системы. Так, в одном из Санкт-Петербургских университетов был создан модуль «ЖИРБИС», предназначенный для построения современного сайта библиотеки. Первая версия этого решения основывалась на стандартном модуле «ИРБИС» для web-поиска — «WEB ИРБИС» и системе управления сайтом Joomla. Там же разрабатывался и оригинальный модуль «ИРБИС-Аналитика», ориентированный на автоматическое заимствование описаний журнальных статей по внутренним протоколам ИРБИС. В омских университетах подготовили и впервые протестировали модуль «Корректор», предназначенный для автоматизированного редактирования каталогов. Усовершенствованные версии этих решений дополнили число официально распространяемых модулей АБИС, а их авторы начали работать в качестве программистов ИРБИС».

В разработанной метапоисковой системе существует возможность выбора электронных библиотек для осуществления поиска. Также в данной системе существует возможность выбора автоматического русско-английского / англо-русского перевода. Перевод осуществляется с помощью <http://translate.google.ru/>. При вводе запроса на русском языке система отправит в русскоязычные электронные библиотеки запрос на русском языке, а в остальные библиотеки — на английском.

Все результаты поиска сгруппированы по электронным библиотекам. Чтобы не загромождать веб-страницу большим количеством информации, пользователю изначально предоставляется набор строк. В каждой строке содержится название электронной библиотеки, количество найденных документов в соответствующей электронной библиотеке и название полей, по которым проводился поиск.

При клике на какой-либо строке раскрывается список результатов поиска по соответствующей электронной библиотеке, содержащий от 10 до 25 документов. Также в этом списке содержится ссылка «Просмотр всех результатов поиска по данному ресурсу», кликнув на которую пользователь перейдет на веб-страницу соответствующей электронной библиотеки со всеми результатами поиска в ней.

Разработанная структура модулей метапоисковой системы допускает расширение системы в случае подключения новых ресурсов. Для этого нужно внести необходимые данные об электронной библиотеке в конфигурационный модуль и добавить в систему модуль-обработчик для нового ресурса, а также написать дополнительную функцию для формирования шаблонов запросов к новому ресурсу.

В случае, если доступ к ранее подключенной электронной библиотеке истек, в конфигурационном модуле электронных библиотек можно установить статус для данной библиотеки «Недоступен» и тогда этот ресурс не будет отображаться на главной странице в разделе доступных ресурсов для поиска.

Надеюсь, что наш журнал читают не только библиотекари, но и разработчики автоматизированных библиотечных систем и создатели онлайн-информационных ресурсов. Поэтому хочу обратить внимание на важность предоставления открытых API к их ресурсам. Это существенно сэкономит ваше время, поскольку часть работы будут выполнять пользователи, и поможет в распространении вашего продукта. А библиотекари должны быть настойчивее, требуя от производителей открытых API. Это поможет интегрировать их продукты в ваши информационные инфраструктуры. ■

По всем интересующим вопросам, а также с предложениями по данному направлению обращайтесь по адресу agpecyshvar@rsl.ru.