

# Современное состояние и перспективы развития проектов Google в библиотечной сфере

www.ubook.ru

## Авторы



**Михаил Владимирович  
ГОНЧАРОВ**

директор Центра развития и поддержки Интернет-технологий ГПНТБ России, к.т.н., доцент кафедры информационных технологий и электронных библиотек МГУКИ



**Кирилл Анатольевич  
КОЛОСОВ**

заведующий отделом Центра развития и поддержки Интернет-технологий ГПНТБ России, к.т.н., доцент кафедры информационных технологий и электронных библиотек МГУКИ

Комплекс порталов и информационных центров Google является признанным лидером и поддерживает поиск по наибольшей в мире коллекции документов, доступных в сети Интернет. За последний год количество проиндексированных документов превысило миллиард. Партнёрами Google по поддержке поиска и предоставлению контента являются тысячи крупнейших компаний, библиотек и информационных центров, включая AOL, Earthlink, Amazon.com, The New York Times, T-Online, Ask.com, Библиотеку Конгресса США, Нью-Йоркскую публичную библиотеку. Google обслуживает 61,8% общемировых поисковых запросов в Интернете, при этом более половины из них поступает от пользователей вне США, где расположена штаб-квартира компании. По данным на конец 2008 г., в компании работают свыше 20200 сотрудников, она имеет более чем 40 офисов в разных странах мира. Интерфейс портала Google поддерживает 112 языков и продублирован на 150 международных доменах сети.

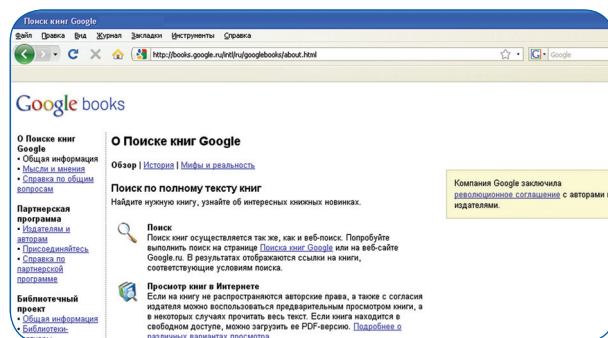
Развитие проектов Google происходит постоянно, и в настоящее время наиболее крупными и популярными по информационному наполнению являются:

- **Web Search** – поиск web-страниц в сети Интернет;
- **News** – поиск новостей из газет и информационных агентств;
- **Book Search** – поиск книг с возможностью частичного доступа к полным текстам;
- **Google Earth** – географический поиск с использованием карт, спутниковых снимков, фотографий объектов и элементами 3D-графики;
- **You Tube** – видеоколлекция, формируемая пользователями сети, с широкими возможностями поиска.

Среди новых проектов заслуживает внимание **Google Ocean**, предназначенный для сбора и представления информации о подводном мире во всех морях и океанах Земли.

Остановимся более подробно на наиболее актуальных для библиотек проектах, уделив особое внимание их развитию в русскоязычной части Интернета.

В рамках проекта **Google Books** реализуется так называемая партнерская программа, при выполнении которой были заключены договора более чем с 20 тыс. издательств и авторов.



Издатели и авторы могут бесплатно предоставлять свои книги для включения их в результаты поиска Google. После получения текстов книг в формате pdf Google добавляет их в поисковый индекс, используемый при обработке запросов пользователей на портале Google Books Search. Каждый результат поиска включает в себя название, автора книги, небольшой фрагмент текста, в котором

цветом выделены условия поиска, и номер страницы фрагмента. Для авторов и издательств это является хорошей дополнительной рекламой, так как Google привлекает пользователей со всего мира, увеличивает посещаемость сайтов издательств, предоставивших информацию. Кроме того, за счет контекстных объявлений, показываемых рядом со страницами книги, издательство и автор могут получать дополнительный доход от Google.

Для представления на портале по партнерской программе было получено более 7 млн книг в электронном формате. Информация предоставляется правообладателем (автором или издателем) в формате pdf. Копия должна соответствовать макету печатного издания (формироваться на основе оригинал-макета верстки). Кроме того, можно просто прислать в Google один печатный экземпляр книги, и сканирование в данном случае проведет Google.

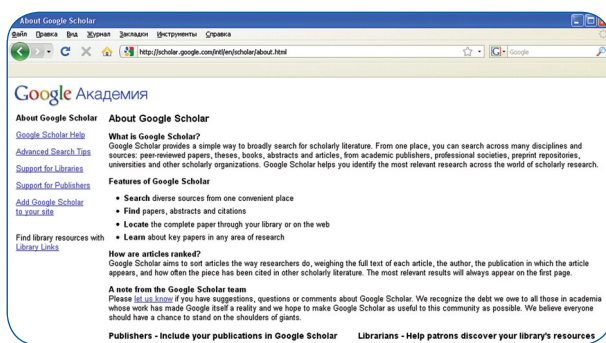
Портал Google Books позволяет искать книги не только в полнотекстовом режиме, но также и по базовым элементам библиографического описания, таким, как автор, заглавие, предметная область. Все книги хранятся на портале в распознанном электронном формате, что позволяет обеспечить полнотекстовый поиск, однако пользователю Интернета доступны лишь фрагменты, он может просмотреть не более 20% издания, исключены возможности печати, сохранения, копирования, а текст показывается в форме последовательности изображений низкого разрешения. Ссылки на магазины позволяют перейти непосредственно к покупке. Издательство или автор, предоставившие книгу, получают детальную информацию о количестве показов книги, рейтинге на портале Google Books Search, а также могут просматривать другую статистику. Правообладатели имеют возможность добавлять на портал или удалять с него любые принадлежащие им книги в любое время.

Книги, изданные до 1923 г., относятся к общественному достоянию, и пользователи портала Google Books имеют возможности для их загрузки в формате pdf, просмотра на мобильных устройствах, копи-

рования и редактирования.

Программные средства портала Google Books имеют возможность привязки библиографических записей, извлекаемых из каталогов библиотек, к режиму предварительного просмотра Google Books на основе программного интерфейса API (Application Programming Interface – интерфейс прикладного программирования).

Проект **Google Scholar** – это удобный интерфейс для поиска учебной и научной литературы с выделением её из огромного массива книг проекта Google Books. Этот проект предусматривает две дополнительных возможности для библиотек: Library Links и Library Search.



**Library Links** – это возможность предоставить пользователю информацию о том, что найденная им статья или книга доступны в электронном формате в конкретной библиотеке. Проект основан на использовании технологии OpenURL (специальный формат URL, включающий метаданные). Гиперссылка, отсылающая пользователя к полному тексту статьи, показывается ему в том случае, если в сети находится библиотека, имеющая подписку на данное электронное издание. В свою очередь, библиотека может бесплатно подключиться к данному проекту, установив у себя программу Link Resolver (определитель ссылок) и направив в Google список электронных изданий, на которые у нее оформлена подписка. Наиболее распространенными вариантами Link Resolver являются программные реализации таких разработчиков, как Ex Libris, Serial's Solutions, SirsiDynix, Teledan, EBSCO, Innovative, Interfaces, Goldrush, OhloLinks, Openly, Informatics North, Carolina, ScholarSFX.

В настоящее время в проекте участвуют около 2000 библиотек.

Однако и сегодня определенные трудности вызывает понимание отечественными специалистами принципов работы технологии OpenURL. Это обуславливает тот факт, что российские библиотеки представлены в данном проекте весьма скудно.

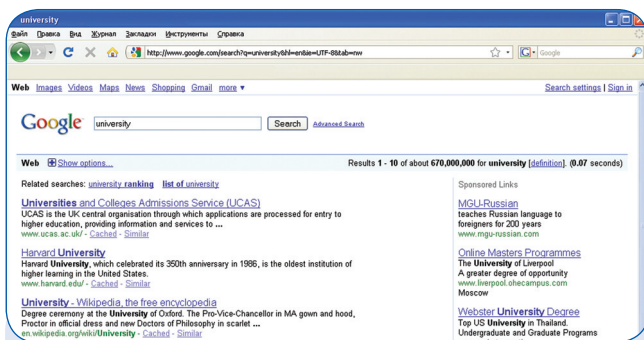
Все процедуры проверки прав доступа и идентификации пользователей выполняются библиотекой и издательством на основе анализа IP-адреса или же с использованием более сложной процедуры. Как правило, при оформлении подписки на электронные версии изданий библиотека заполняет регистрационную форму, где указывает диапазон IP-адресов компьютеров, с которых будет осуществляться доступ к ресурсам. Пользователи, работающие в помещениях библиотеки, получают доступ к Интернету через проxy-сервер или шлюз библиотеки. Любая попытка обращения к подписным электронным ресурсам извне библиотеки будет пресекаться сервером издательства или подписного агентства. Кроме того, число одновременных сеансов может быть ограничено договором с издательством.

Издательство или подписное агентство может предоставить библиотеке для администрирования доступ к подписным ресурсам на основе проверки имени пользователя и пароля, однако такой способ не рекомендуется для использования читателями.

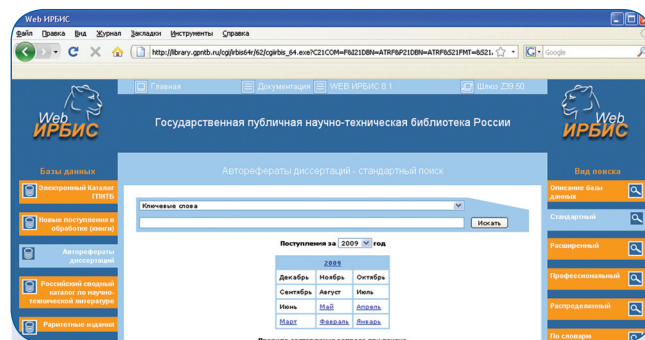
**Library Search** предоставляет возможность найти ближайшую библиотеку, где есть книга, которую пользователь нашел на портале Google Scholar. В проекте участвуют более 25 сводных каталогов, среди которых наибольшим по объему является WorldCat. Пользователь переадресовывается к выбранному сводному каталогу, который по IP-адресу компьютера определяет местонахождение ближайшей библиотеки.

В настоящее время для русскоязычной литературы точкой входа является поиск на портале Российского сводного каталога по научно-технической литературе (Search GPNTB Catalog). Автоматизированная система Российского Сводного каталога по НТЛ функционирует с 1980 г. и содержит более 720 тыс. библиографических





Пример вывода результатов поиска на портале Google Scholar с гиперссылкой на запись в Российском сводном каталоге по НТЛ



Пример вывода результатов поиска после перехода пользователя на сервер Российского сводного каталога по НТЛ в ГПНТБ России.

записей из фондов 700 библиотек России и стран СНГ. Внутренний формат сводного каталога базируется на международном формате UNIMARC и отвечает требованиям многих автоматизированных систем, используемых библиотеками. Экспорт и импорт записей возможен в форматах RUSMARC, UNIMARC, MARC21 (в представлении ISO-2709 или построчном текстовом формате). Помимо основного электронного формата, данные, загружаемые в сводный каталог, принимаются в ГПНТБ России и в текстовом виде. Важный момент: система никогда не претендовала на то, чтобы стать эталоном библиографической записи, так как основная ее цель – информирование о местонахождении издания, и во всех описаниях она представлена как адресно-библиографическая система. Дальнейшее развитие технологии предполагает создание и поддержку новых информационных сервисов, таких как ведение распределенного адресно-библиографического банка данных (электронные каталоги), полнотекстовых баз данных (электронные библиотеки), обеспечение возможностей удаленного заказа документов по МБА и ЭДД, а также сервисов корпоративной и ретроспективной каталогизации.

В настоящее время пользователи сводного каталога имеют возможность получить необходимые им издания, воспользовавшись службой удаленного заказа документов по МБА и ЭДД, переходя по гиперссылке на web-страницу библиотеки – держателя документа. Реализована возможность связывания библиографических записей сводного каталога с полными текстами на сайтах издательств и подписных агентств, что позволяет давать читателям доступ к документам непосредственно из читальных

залов библиотек. Сразу после перехода по гиперссылке пользователь видит библиографическую запись с сервера ГПНТБ России, в которой содержатся сиглы (уникальные идентификаторы) библиотек – держателей документа.

Логическим развитием принципа последовательной навигации в web-среде при поиске ресурса является использование сиглы библиотеки в качестве гиперссылки, отсылающей к соответствующей записи в каталоге (ОРАС) данной библиотеки. Дальнейшая навигация (возможность заказа документа на кафедру, по ЭДД и т.п.) будет происходить уже средствами сайта библиотеки – держателя документа.

Техническое решение для реализации предложенного варианта базируется на применении технологии Z39.50 и протокола SRU/SRW, а также на использовании единого коммуникативного формата MARCXML, дополненного полями, содержащими гиперссылки на записи в каталоге библиотеки – держателя документа, в соответствии с технической спецификацией компании Google. Библиотеки, участвующие в новой реализации Российского сводного каталога по НТЛ, предоставляют доступ к своим библиографическим записям через серверы Z39.50 в коммуникативном формате RUSMARC или MARC21. Одно из полей коммуникативного формата будет содержать гиперссылку, указывающую на эту же библиографическую запись на web-интерфейсе ОРАС. Для автоматического обслуживания запросов по протоколу SRU/SRW можно использовать свободно распространяемый программный пакет YAZ Proxu. Протокол SRU/SRW позволяет представлять данные, полученные с сервера Z39.50 в формате XML записи в соответствии с указанной схемой метаданных (MARCXML или

DublinCore).

В ГПНТБ России предлагается создать программную подсистему автоматического сбора метаданных, основу которой составит текущий массив данных Российского сводного каталога по НТЛ. Преобразование сигл библиотек в гиперссылки будет осуществляться автоматически в результате последовательного опроса серверов Z39.50 (по протоколу SRU) библиотек, желающих участвовать в проекте. Программа опроса будет производить поиск в каталоге удаленной библиотеки (например, по полю ISBN записи сводного каталога), извлекать обратную гиперссылку, содержащуюся в соответствующем поле, и подставлять её в массив MARCXML, используемый для индексации Google на сервере ГПНТБ России.

Кроме того, данная технология позволит в автоматическом режиме пополнять записи Российского сводного каталога по НТЛ. Программа опроса может обращаться к удаленным серверам Z39.50 библиотек, участвующих в проекте, с поисковыми запросами, например, по году издания. Если запись, полученная от сервера, содержится в Российском сводном каталоге по НТЛ, то будет осуществляться лишь преобразование сиглы в обратную гиперссылку, как это было описано выше. Если же запись новая, то она будет автоматически полностью добавлена в сводный каталог и массив, индексируемый Google. Использование в предложенной технологии первичных записей в формате MARCXML, формируемых библиотеками – держателями изданий, позволит поддерживать массив записей на сервере ГПНТБ России, используемый для индексации Google, в актуальном состоянии без необходимости преобразований данных, связанных с возможными искажениями при их обработке.