

**Поиск книг в программе Google. Взгляд историка**  
**Google Book Search. A Historian's View**  
**Пошук книг за допомогою серверу Google. Погляд історика**

*Теренс Эммонс*  
*Стэнфордский университет, Стэнфорд, Калифорния, США*

*Terence Emmons*  
*Stanford University, Stanford, California, USA*

*Теренс Еммонс*  
*Стенфордський університет, Стенфорд, Каліфорнія, США*

Представлен особый взгляд автора-историка на новый проект Google – «Google Library Project». Основа проекта – договор с крупнейшими библиотеками (в том числе Оксфордского, Гарвардского, Станфордского и Мичиганского университетов, а также Публичной библиотекой Нью-Йорка) о переводе книг в цифровой формат для компьютерной обработки в интернете с помощью новой (и пока засекреченной) технологии сканирования.

The special opinion of the author-historian about the new project by Google («Google Library Project») is presented. The basic idea of the Project is concluding contracts with the major libraries (including Oxford, Harvard, Stanford and Michigan University Libraries, as well as New York Public Library) on digitization of books from their collections for further computer processing in the Internet using new (blackout at this time) scanning technology.

Представлено особливий погляд автора-історика на новий проект Google – «Google Library Project». Основною проекту є угоди з найбільшими бібліотеками (включно з бібліотеками Оксфордського, Гарвардського, Стенфордського та Мічиганського університетів, а також Публічною бібліотекою Нью-Йорка) про переведення книг у цифровий формат для комп'ютерної обробки в Інтернеті за допомогою нової (і поки таємної) технології сканування.

14 декабря 2004 г. фирма Google сделала сенсационное сообщение о договоре с крупнейшими библиотеками, в том числе с библиотеками Оксфордского, Гарвардского, Станфордского и Мичиганского университетов, а также Публичной библиотекой г. Нью-Йорк. В сообщении говорилось о том, что миллионы и миллионы книг будут переводиться в цифровой формат для компьютерной обработки в интернете с помощью новой (и пока засекреченной) технологии сканирования. Каждый договор имеет свою специфику, например, Станфордская и Мичиганская библиотеки условились включить в эту программу свои коллекции целиком! Срок, отведенный для сканирования этих многомиллионных коллекций – 6 лет! Называется проект – «Google Library Project».<sup>1</sup>

Программу, конечно, нельзя назвать беспрецедентной. Многие библиотеки уже начали частично переводить в цифровой формат свои коллекции, в основном, старинные и редкие издания, для обеспечения их сохранности (т.е. продолжили микрофильмирование с помощью новой технологии). Amazon.com, самая крупная фирма по продаже книг в Интернете, уже несколько лет использует технологию, позволяющую пользователям (по предварительному соглашению с издательством в каждом отдельном случае) увидеть обложку, титульный лист, содержание, несколько страниц текста предлагаемого издания, и т.д. («Search Inside the Book»). «Проект Гутенберг» был начат задолго до развертывания интернета, и на сегодняшний день этот своеобразный склад «электронных книг» («ebooks») содержит много тысяч полных текстов, в основном, англоязычных книг, не ограниченных авторским правом, и т.п.<sup>2</sup> В 2003 г. Google запустил программу под названием «Print

<sup>1</sup> <http://news.bbc.co.uk/go/pr/ft/-/1/hi/technology/4094271.stm> 14.12.2004.

[http://www.google.com/press/pressrel/print\\_library.html](http://www.google.com/press/pressrel/print_library.html) 14.12.2004.

<http://searchenginewatch.com/searchday/article.php/3447411>. См. также сайты названных библиотек, где даются детали договоров.

<sup>2</sup> «Project Gutenberg Mission Statement» <http://www.gutenberg.org/about/faq0>.

Program» или «Publishers' Program», или же «Books Partner Program,» похожую на ту, что Amazon.com предлагает своим читателям. Эта программа по соглашению с издателем каждой отдельной книги позволяет пользователям ознакомиться с содержанием книг. Изображение страницы на экране сопровождается сопутствующими рекламами, помещенными в правом поле с помощью знаменитой технологии Google («AdSense contextual listings»). Доходы от рекламы распределяются между фирмой и издателями.

Новая программа Google, видимо, строится на основе Print Program. В октябре 2004 г. фирма объявила, что издательства могут продавать полные цифровые тексты своих изданий через Print Program,<sup>3</sup> а уже в декабре последовало объявление о новой «библиотечной» программе. Здесь следует отметить, что книжная информация, получаемая по «Print Program» или по новой программе, не включена в общую базу данных Google (так называемый «web index»), с помощью которой выполняется обычный поиск. В отличие от Print Program библиотечная программа денег не приносит, реклама на экране отсутствует, и пользователь видит только книжную информацию, в основном факсимильную. Google утверждает, что цель проекта—небывалое распространение сведений о книгах, которое принесет пользу всем, включая авторов и издательства, а также беспримерное расширение базы данных для поиска информации обо всем на свете. При этом полный текст разрешается просматривать только в старых книгах, не ограниченных авторским правом (изданных до 1922 г. по американскому праву), а в остальных случаях (за исключением конкретных объектов договоренности по Print Program) можно увидеть только искомые «ключевые слова» и несколько строк, иногда несколько примыкающих страниц, плюс обложку, содержание, указатели, библиографическое описание, и (что очень важно) разрешается поиск других терминов в тексте данной книги. Печатать тексты не разрешается («printing function disabled»).

Таким образом, Google утверждает, что копирование им библиотечных книг без предварительного разрешения автора или издательства не нарушает их права, потому что книги будут использоваться исключительно с поисковой целью и дадут пользователю возможность ознакомиться только с небольшими отрывками текста (так называемые «snippets»). Иначе говоря, программа делает то же, что без разрешения автора или издателя делает любой писатель/исследователь в процессе подготовки своей книги или статьи. Программа не только не нарушает авторские права и не наносит ущерб продаже книг, а напротив, утверждает Google, делает информацию о книгах более доступной, а с помощью сообщения в левом поле о том, где можно приобрести данную книгу или как найти ее в ближайшей библиотеке, стимулирует более широкую продажу книг и их использование в целом.<sup>4</sup>

Перечисленные библиотеки включились в программу по разным соображениям; общим знаменателем для принятия решения является перспектива сохранения их коллекций. Цифровые копии отсканированных книг остаются собственностью библиотек. Главный библиотекарь Гарварда отметил, что на работу, которую Google собирается сделать за шесть лет, библиотеке с существующей у нее технологией понадобилась бы тысяча лет. Библиотеки руководствовались и другими соображениями, например, их привлекает возможность обеспечить пользователям более широкий доступ к коллекциям. Станфордские библиотекари явно рассчитывают на возникновение новых способов обработки оцифрованной информации; таким образом, база данных для применения новой технологии будет уже подготовлена.

Но если библиотечный мир встретил новый проект Google с энтузиазмом, то в издательском мире и, в какой-то мере среди авторов, отношение к нему было иным. Разные издательские группы, как Association of American Publishers (AAP), ассоциация университетских издательств (Association of American University Presses), авторская гильдия (Authors Guild) и целый ряд авторских группировок и отдельных авторов сразу же подали на Google в суд, обвинив фирму в нарушении авторских прав. Все они утверждают, что только авторы или издатели (в зависимости от того, кто владеет авторским правом) имеют право разрешать копирование книг, находящихся под защитой этого

---

<sup>3</sup> «Google Print Opens Widely to Publishers» <http://searchenginewatch.com/searchday/article.php/3417941>.

<sup>4</sup> «Information for publishers and authors about the Library Project» [http://print.google.com/googleprint/publisher\\_library.html](http://print.google.com/googleprint/publisher_library.html). «Library Project—An enhanced card catalog of the world's books» <http://print.google.com/googprint/library.html>.

закона. Их не убеждают утверждения Google, что включение книг в программу будет выгодно всем, что Google этой программой деньги не зарабатывает, и другие аргументы.<sup>5</sup> Существуют подозрения, что так или иначе Google получит прибыль за их счет. Притом, как отмечает директор одного университетского издательства, электронные копии книг, которые остаются во владении библиотек, принимающих участие в проекте, полностью выйдут из-под контроля Google, то есть ограничения, соблюдаемые Google в своей программе в отношении книг, не распространяются на библиотечные копии книг. Никому не известно, что библиотеки будут делать со своими копиями, и процесс этот никем не контролируется.<sup>6</sup>

Не знаю, разумно ли такое отношение к новой программе Google.<sup>7</sup> Некоторые циники считают, что недовольство со стороны некоторых издательств, вызвано опасением, что библиотечная программа заслонит и постепенно вытеснит Publishers' Program, а они рассчитывали получить значительные доходы от ее использования. Фактически обе программы пока сосуществуют: в результате поиска информации о каком-то предмете можно получить или страницу библиотечной книги без рекламы, или информацию, предоставляемую Publisher's Program, с рекламой. Оба типа информации пользователь получает на одном сайте: Google Book Search.<sup>8</sup>

Google ответил на град судебных дел временной приостановкой в августе 2005 г. сканирования книг, защищенных авторским правом (практически всех книг, изданных после 1922 г.) и предложил издательствам составить и прислать списки тех книг, которые не следует сканировать в рамках проекта.<sup>9</sup> Издательствам было предложено включить в Publishers' Program книги, подлежащие сканированию по библиотечной программе (см. выше о сосуществовании двух программ). В то же время, Google не отступил от своей позиции относительно законности массового сканирования всяких книг и создания базы данных с чисто поисковой целью. В связи с этим он запустил энергичную кампанию в средствах массовой информации, и объявил, что возобновит неограниченное сканирование библиотечных книг 1 ноября 2005 г.

Так Google и сделал, по крайней мере формально, несмотря на то, что судебные иски продолжают.<sup>10</sup> «Формально», потому что библиотеки, участвующие в проекте, проявляют осторожность, ожидая как решится юридический статус проекта. Некоторые из них с самого начала согласились пока только на сканирование старых книг, а те, кто подписал соглашение о сканировании их коллекций целиком, выжидают. По крайней мере, Станфорд пока не разрешает сканировать издания, охраняемые авторским правом (личное сообщение директора Станфордских библиотек).

Теперь перейдем к вопросу о том, насколько новая программа Google полезна для историков.

---

<sup>5</sup> «Publishers Sue Google Over Plans to Digitize Books»

<http://publishers.org/press/releases.cfm?PressReleaseArticleID=292>. 19.10.2005. «Authors Guild Sues Google, Citing «Massive Copyright Infringement. September 20, 2005 Press Release» [http://www.authorsguild.org/news/sues\\_google\\_citing.htm](http://www.authorsguild.org/news/sues_google_citing.htm). «University-Press Group Raises Questions About Google's Library-Scanning Project» <http://chronicle.com/cgi2-bin/printable.cgi?article>. 23.05.2005. «The Google Print Controversy: A Bibliography» <http://www.escholarlypub.com/digitaloans/2005/10/25/the-google>.

<sup>6</sup> «Fair Use in Theory and Practice: Reflections on Its History and the Google Case,» Sanford G. Thatcher, Director, Penn State University Press. Prepared for the panel on Dark Archives and Celestial Card Catalogues: Google Print and the Future of Fair Use at the NACUA conference on The Wired University: Legal Issues at the Copyright, Computer Law, and Internet Intersection, Arlington, VA. November 10, 2005.

<sup>7</sup> Jonathan Band, «The Google Library Project: the Copyright Debate» OITP [Office for Information Technology Policy, American Library Association]. January 2006.

<sup>8</sup> «Google Library Used to Undercut Google Print» [http://www.techdirt.com/articles/20050617/0941208\\_F.shtml](http://www.techdirt.com/articles/20050617/0941208_F.shtml).

<sup>9</sup> Margaret Kane, «Google pauses library project» [http://news.com.com/2102-1025\\_3-5830035.html](http://news.com.com/2102-1025_3-5830035.html). Association of American University Presses. Issues in Scholarly Publishing, «Google's New Opt-Out Policy» <http://aaupnet.org/aboutup/issues/81905.html>.

<sup>10</sup> Mark Chillingworth, «Google library scanners back in action,» Information World Review 01 Nov 2005 <http://www.iwr.co.uk/articless/print/2145276>. Nancy Gohring «Google Resumes Book-Scanning Project. Controversy, lawsuits remain over building a database of copyrighted works,» PC WORLD, November 02, 2005 <http://www.pcworld.com/resource/printable/article/0.aid.123374.00.asp>.

Здесь, конечно, следует иметь в виду, что процесс сканирования литературы и создания базы данных еще не завершен, а в отношении книг на иностранных (т.е. не англоязычных) языках, видимо, только-только начинается. Этот вопрос, между прочим, пока не вполне ясен: например, Станфордские библиотекари говорят, что для них язык не является критерием в отборе книг для сканирования, но предполагают, что сам Google пока воздерживается от массовой обработки иностранной литературы «по чисто техническим причинам». Тем не менее, количество книг, введенных в базу данных непосредственно из библиотечных хранилищ или с помощью Publishers' Program, является весьма внушительным. Это во-первых.

Во-вторых, с возникновением новых поисковых технологий открываются возможности для новых подходов историков к прошлому. Например, поиск ключевых слов позволяет на уровне книг использовать методы анализа содержания («content analysis») в изучении общественного мнения, которые до сих пор использовались при анализе газетного материала. (Здесь следует отметить, что Google Book Search никак не помогает историкам решить проблемы, связанные с поиском библиографических справок в отношении периодической литературы).

Итак, ограничусь результатами нескольких попыток одного историка довольно традиционного профиля использовать Google Book Search. Первый опыт такой. Я недавно занимался подготовкой к печати незаконченной рукописи своего покойного учителя Мартина Малиа (Martin Malia).<sup>11</sup> В одном месте автор приводит без ссылки на источник две строки английского поэта Вильяма Вордсворта о Великой французской революции:

**Bliss was it in that dawn to be alive**

**But to be young was very heaven**

Ввожу эти строки в Google Book Search, Advanced Search (где можно направить поиск—точно как в онлайн-каталогах разных библиотек—на работы одного автора, на определенное название книги, по датам издания, с ограничением количества результатов, и т.д.), в данном случае с указанием имени автора, и сразу получаю около 10 ссылок. Из них две ссылки давали тот отрывок текста поэмы «The Prelude» где находятся эти строки. Одна ссылка поступила в базу данных через библиотечную программу (так и написано вверху: «from the Google Books Library Project»), разрешающую распечатку страницы с известными строками, а также просмотр книги целиком (речь идет о книге, изданной до 1922 г.). Вторая ссылка поступила в систему через Publishers' Program («Sample pages from the Google Books Partner Program»), в данном случае собрание стихотворений Вордсворта, опубликованное Oxford University Press в 2000 году. Представлена страница текста с выделенными искомыми строками и одна-две прилегающие страницы, но распечатка не разрешается. Первая ссылка—без рекламы, и она содержит только информацию о том, где можно приобрести книгу или как найти ее в библиотеке. Вторая ссылка представляет ту же информацию о том, где можно купить книгу, но без библиотечного поиска. И что очень интересно, она дает возможность увидеть обложку, содержание, указатели, но не титульный лист. Информация о титульном листе дается в конце в нефаксимильной форме. Есть и рекламы, в данном случае, не очень подходящие («Bee-Alive Royal Jelly—т.е. косметика; «Bliss-Official Site»—тоже, почему-то, косметика; и «Discount Items. Products and info about things you need»—реклама ebay.com). Но и за это спасибо—разыскал источник цитаты.

Второй случай. Мой автор часто цитирует известные слова Маркса о том, что революции являются локомотивами истории (по-английски «locomotives of history»). Печатаю фразу, опять же с указанием автора, в Google Book Search. Результат (за 0.03 секунды): 4 ссылки, из которых всего в одной указано сочинение самого Маркса (остальные—комментарии): «Классовая борьба во Франции: от февральской революции к Парижской коммуне». Текст известной работы Маркса в английском переводе приводится по «Google Book Partners Program»—с такими же ограничениями доступа, как и в предыдущем случае, и опять же с не очень логичными рекламами (материалы для миниатюрных поездов [2] (!) и опять же ebay).

---

<sup>11</sup> History's Locomotives: Revolutions and the Making of the Modern World (New Haven: Yale University Press, 2006 (in press)).

Но цитата из Маркса все-таки нашлась! Единственный ли это случай употребления Марксом железнодорожной метафоры? Трудно сказать, пока не все тексты Маркса введены в программу. Но вряд ли прав мой знакомый, утверждавший, что данная фраза встречается в тексте Коммунистического манифеста. Манифест ведь уже введен в программу.

Итак, дважды обратившись к программе Google с такой типичной для историка или редактора задачей, дважды получил нужную справку.

Помню, что встретил объявление Google об их новой программе довольно скептически: что нового она может добавить в фундаментальное для историка дело поиска библиографических справок, чего не дают уже существующие каталоги крупнейших библиотек мира или сборные каталоги вроде RLG, OCLC World Cat., и т.д.? В связи с этим я решил провести маленький эксперимент. Начнем с нескольких общих замечаний.

Как выглядят результаты Book Search по сравнению с перечисленными каталогами для исследователя, желающего получить библиографическую справку на определенную тему? Здесь, с моей точки зрения, следует отметить два момента:

1. С помощью Google можно ознакомиться с содержанием книг (т.е. с перечнем глав), а также с указателями. Каталоги дают только описание книги, сделанное в процессе каталогизации с помощью того или другого справочника по систематизации. (Отдельный предмет поиска составляют полные тексты, доступные онлайн).

2. Если я правильно понимаю (не будучи профессиональным библиотекарем), тематические поиски в библиотечных каталогах проводятся на базе «ключевых слов.» Тематические поиски библиотечных каталогов по ключевым словам базируются *faute de mieux* на названиях и описаниях книг, каталогизированных с использованием строго контролируемого словаря (в Америке этот словарь зафиксирован в «LC Subject Headings»). Если указанных слов нет ни в названии книги, ни в ее описании, то эта книга в результате тематического поиска выявлена не будет. Google Book Search тоже построен на ключевых словах, но Book Search проводит поиск всего содержания книги, а не только названия или описания.

Сравним результаты поиска по одной и той исторической теме с использованием сводного каталога RLG и Google Book Search. Для этого эксперимента я выбрал событие, которому до сих пор уделяется очень много внимания в американской историографии, а именно «The Dred Scott Decision» 1857 г., пресловутое решение Верховного суда, по которому рабы и потомки рабов, несмотря на место жительства (включая так называемые «свободные штаты» Севера) не могли быть гражданами Соединенных Штатов, и соответственно лишались судебной защиты и других прав. Многие историки считают, что это решение Верховного суда обусловило неизбежность гражданской войны в США. В результате и «Dred Scott Decision», и «Dred Scott» имя человека, который начал дело, прошедшее с помощью прогрессивных северных адвокатов все судебные инстанции вплоть до Верховного Суда, были введены в политическую полемику и стали широко используемыми метафорами.

Начнем с поиска по словам «Dred Scott» («keyword search») в RLG. Результат: 348 названий книг, многие из них со словами «Dred Scott» или «Dred Scott Decision» в заглавии книг, но многие заглавия не содержат этих слов, они подобраны, должно быть, по описаниям каталогизаторов. (Интересно, что поиск по заглавиям дает почти столько же названий). Просмотрев этот список, трудно утверждать, что он построен по степени релевантности. Книга моего покойного коллеги по Станфорду, Дона Ференбахера (Don Fehrenbacher), за которую он получил Пулитцеровскую премию в 1978 г.<sup>12</sup>, пожалуй, самая известная и авторитетная книга, посвященная этому делу, стоит на 327-ом месте (поиск сделан без указания даты публикации). Между прочим, поиск по рубрике «автор» «Don Fehrenbacher» дает 78 названий. Проведя поиск по ключевым словам «Dred Scott» по каталогу Станфордской библиотеки, я получил вместо 348 названий всего 52.

Теперь посмотрим, что даст поиск в программе Google Book Search по тем же параметрам. Поиск по заглавиям книг со словами «Dred Scott» выявляет всего-навсего 7 названий, в том числе 2 книги, написанные Доном Ференбахером. Но

---

<sup>12</sup> Don Edward Fehrenbacher, *The Dred Scott Case: its Significance in American Law and Politics* (New York: Oxford University Press, 1978).

простой поиск по всем книгам слов «Dred Scott» выявляет уже 278 книг с 25,900 страницами с упоминанием этих слов. Книги расположены в списке по критерию упоминаемости искомых слов. Так, в упомянутой книге Ференбахера, полностью посвященной этой теме, второй в списке, имя Dred Scott упоминается 768 раз (что равняется общему количеству страниц в книге). В книге No. 100, Dred Scott упоминается 9 раз; в No. 190–6 раз, а в последней по списку, No. 278 – всего 3 раза. Поиск по имени автора (Fehrenbacher) дает всего 3 названия книг.

Я сказал выше, что имя Dred Scott стало нарицательным. Это подтверждают результаты поиска этого имени не по Book Search Google, а по Web Search: «Dred Scott Decision» выявляет «приблизительно» 510,000 упоминаний во всевозможных источниках. Уберем слово «decision» и получается «приблизительно» 1,210,000 упоминаний, из которых «приблизительно» 1,080,000 в публикациях, вышедших в течение последнего календарного года.

Общие выводы. Для стандартной библиографической справки или составления списка литературы на определенную тему каталоги (да и специальные библиографии) программой Book Search не заменишь. Она далеко не завершена – как свидетельствуют приведенные выше цифры о поисках по названиям книг и по автору – и речь здесь шла только о книгах на английском языке. Для поиска иностранной литературы мы должны пока почти полностью опираться на каталоги. В то же время Google, в отличие от каталогов, позволяет исследователю заглянуть внутрь книги, увидеть содержание и указатели. Трудно переоценить практическое значение этого качества для исследователя. Со временем Google Book Search может стать самым универсальным и соответственно самым удобным сводным каталогом в мире. А пока и Google Book Search, и библиотечные каталоги необходимы историку в поиске библиографических справок. Они дополняют друг друга. О других потенциальных возможностях Google Book Search, связанных с расширением базы книжных данных и разработкой новых поисковых технологий, вероятно, еще рано говорить.