

КОММЕРЧЕСКИЕ ПОЛНОТЕКСТОВЫЕ БАЗЫ ДАННЫХ И ИХ ИСПОЛЬЗОВАНИЕ В БИБЛИОТЕКАХ

Цифровая среда, обеспечивающая быстрый и надежный доступ к информационным ресурсам из любой точки планеты, явилась катализатором развития всевозможных форм электронного бизнеса. Одними из первых оценили преимущества Интернет компании, специализирующиеся на производстве и продаже всевозможных баз данных. Созданием подобных баз с начала 80-х годов занималось множество зарубежных корпораций, главным продуктом которых была информация. Существовали фактографические базы данных, содержащие фактические сведения (прежде всего статистику), библиографическую информацию (сведения о документах) и полнотекстовые (полные тексты книг и статей из газет, журналов и сборников). Среди наиболее известных производителей и поставщиков баз данных в «доинтернетовский» период выделялись LEXIS/NEXIS, Dialog, Silver Platter, EBSCO Information Services, STN International, H.W.Wilson, UMI (ныне ProQuest). Информационные продукты доставлялись пользователям на магнитных лентах, посредством модемной связи, а со второй половины 80-х годов – на CD-ROM.

С распространением Интернет информационный бизнес приобрел невиданные до сего времени масштабы. К уже существовавшим гигантам информационной индустрии добавились молодые компании, среди которых наиболее заметны netLibrary и Northern Light. Для поставщиков информационных продуктов Сеть явилась идеальной средой ведения бизнеса: информационные корпорации обрели возможность вести круглосуточное обслуживание клиентов вне зависимости от их местонахождения. Потенциальный рынок враз расширился до пределов всего земного шара. При этом появилась возможность оказывать информационные услуги клиентам напрямую, обходясь в большинстве случаев без привлечения посредников, в лице которых ранее выступали библиотеки.

Из всего перечня представленных в Сети коммерческих информационных продуктов для профессиональной информационной деятельности наибольший интерес представляют полнотекстовые базы данных. В них содержатся тексты книг, статей из журналов, газет и сборников, сообщения информационных агентств, аналитические отчеты различных учреждений и другие документы. Для традиционных библиотек получение доступа к этим базам в значительной степени меняет подходы к информационной работе в целом, так как позволяет читателям обходиться без обращения к печатным оригиналам изданий, и, следовательно, избавляет потенциальных клиентов от необходимости физического посещения библиотек.

Число и отраслевой спектр коммерческих полнотекстовых электронных собраний постоянно расширяется. Если на заре этого вида

бизнеса в 60-е годы их создание начиналось с правовой и экономической областей, в 80-90-е года были освоены естественные, точные и наиболее динамично развивающиеся гуманитарные дисциплины, то в настоящее время очередь дошла до коллекций редких и старопечатных книг, полных собраний сочинений античных и средневековых авторов, поэтических антологий и подобных этому материалов.

Компании, непосредственно предоставляющие доступ к полнотекстовым собраниям, имеют в арсенале сразу несколько десятков или даже сотен баз данных от разных производителей. Их число нестабильно, хотя наблюдается зримая тенденция к увеличению общего количества. В зависимости от пользовательского спроса и, следовательно, экономической оправданности, различные базы данных появляются или удаляются из доступа. Зачастую одну и ту же известную базу данных можно встретить у различных поставщиков. В данном случае компании конкурируют друг с другом по уровню комфортности доступа и, конечно же, в области цен, что, безусловно, выгодно конечным пользователям.

Наполнение полнотекстовых баз данных производится двумя способами. Первый заключается в сканировании печатных оригиналов и получении электронных копий документов, выполненных в большинстве случаев в формате PDF. Эти документы воспроизводятся (читаются) с использованием бесплатно распространяемой программы Adobe Acrobat, которую можно без труда получить из Интернет или найти на большинстве выпускаемых в настоящее время компакт-дисков (в этом случае диск маркируется логотипом Adobe Acrobat). Сканирование печатных оригиналов применяется при оцифровывании существующих тематически и логически законченных собраний, хранящихся, как правило, в фондах крупнейших библиотек или архивов. При этом масштабы и темпы оцифровки документальных массивов в рамках коммерческих проектов в разных странах дают все основания полагать, что уже в ближайшие годы будет оцифровано большинство значимых для человечества материалов, созданных в «доцифровую эру».

Второй метод - покупка электронных копий книг, газет или журналов непосредственно в издательствах. Таким образом приобретается большинство современных изданий. По предварительному договору издательства передают электронную версию документа (чаще всего готовый оригинал-макет) поставщику и получают определенный процент от средств, полученных за обращение к поставленным файлам. Загрузка документов в систему осуществляется, как правило, в момент опубликования печатного оригинала или даже ранее. После физической загрузки в базу информационный массив индексируется, после чего электронные документы становятся доступными для поиска и выгрузки. Подобный способ «комплектования» применяется при работе с современными периодическими изданиями и сообщениями агентств новостей.

Сочетание данных методов комплектования в ближайшей перспективе приведет к тому, что для пользователей станет принципиально возможным

получить доступ к любым источникам - от германских старопечатных книг XVI века до заметки в завтрашнем номере японской газеты.

Профессиональные цифровые библиотеки, в отличие от бесплатных коллекций, отличается намного более четкая политика в отборе источников, высокая степень полноты и оперативность актуализации материалов. Грамматические ошибки крайне редки, за исключением случаев, когда документы сохраняются не в PDF, а в HTML-формате. Как и большинство бизнес-проектов, полнотекстовые базы данных, обеспечивают для клиентов высокий уровень сервиса. Поисковый механизм позволяет осуществлять многоаспектный поиск с возможностью сочетания данных из разных полей. Разыскание может осуществляться по отдельным словам, словосочетаниям и точным фразам. Результаты поиска выдаются в виде списка библиографических записей с указанием всех необходимых элементов. Существует возможность формирования из общего перечня списка релевантных документов.

Сами полные тексты предстают перед пользователем в виде HTML-документов или PDF-файлов. Особенностью последних является способность представлять документ именно в том виде, как он существует в печатной копии, с сохранением колонок, таблиц, иллюстраций и т.д. В тоже время все содержание такого документа может быть доступно для поиска (проиндексировано). Источники в PDF можно не только просматривать (читать) в онлайн-режиме, но и сохранять на собственном компьютере для дальнейшей работы.

Пользование коммерческими базами осуществляется по предварительно подписке, которая оформляется, как правило, на один год. Технически доступ осуществляется по предварительно выданным имени пользователя и паролю или по IP-адресу. Последний способ наиболее выгоден для организаций, в частности библиотек, поскольку позволяет предоставлять пользование такими базами данных с любых компьютеров, расположенных в учреждении (находящихся за одним IP-адресом).

В настоящее время стоимость доступа к полнотекстовым базам данных весьма высока. Серьезные затраты на сканирование и распознавание текста, описание и индексирование документов, приобретение лицензий от владельцев авторских прав делают себестоимость конечного продукта весьма высокой. Сравнительно малое число подписчиков заставляет владельцев электронных собраний держать высокие цены, дабы окупить собственные затраты и получить прибыль (при существенном повышении числа подписчиков стоимость доступа для каждого отдельного пользователя может быть значительно снижена). Однако уже сегодня возможность получения огромного массива данных в любой точке земного шара, где есть доступ к Интернет, привлекает все более значительное число пользователей, среди которых безусловное лидерство держат крупные корпорации, деятельность которых основывается на анализе информации. Высокая стоимость доступа для остальных категорий пользователей несколько смягчается позицией традиционных библиотек, которые в ряде случаев предпочитают подписку на

полнотекстовые базы данных, непосредственному приобретению книг и журналов в собственный фонд.

Кроме того, практически все коммерческие базы данных позволяют в той или иной мере воспользоваться своими ресурсами без оформления платной подписки. Некоторые базы можно весьма эффективно применять в качестве библиографических источников, без возможности получать полные тексты документов. Типичным примером является сервис Library Reference Center (<http://search.epnet.com/reflogin.asp>) компании EBSCO, позволяющий осуществлять библиографический поиск в базе данных статей из более чем 30 ведущих англоязычных библиотечных журналов. Другие полнотекстовые библиотеки открывают доступ для всех желающих в «непиковое» время или на определенный период – в виде «ресурса месяца» и подобных форм. Внимательное изучение характеристик представленных источников поможет выявить те из них, которые обладают дополнительными ценными свойствами. Классическим примером служит предоставление компанией ProQuest доступа к авторефератам диссертаций, защищенных в США, Канаде и Тихоокеанском регионе за последние два года (<http://wwwlib.umi.com/dissertations>). Эта база позволяет выполнять многоаспектный поиск, просматривать полные тексты авторефератов в виде сканированных копий и выводить их на печать. Наконец последним и практически уникальным «ключом» к ресурсам коммерческих цифровых собраний при ограниченности бюджета является оформление бесплатного «пробного доступа» (free trail). Естественно, что такие методы мало годятся для систематического использования ресурсов, однако крайне полезны для детального ознакомления с возможностями различных баз данных.

В последнее пятилетие среди информационных компаний прокатилась череда слияний и поглощений, свидетельствующая о высокой интенсивности развития данного сектора экономики. Информационные компании постоянно расширяют свой бизнес. Помимо простого предоставления доступа к коллекциям, владельцы полнотекстовых баз данных создают дополнительные сервисы, для сфер, где использование электронных библиотек наиболее эффективно. Таковы, например, проект XanEdu компании ProQuest, предназначенный для университетского сообщества, ReferenceLink компании Thomson Corporation, рассчитанный на библиотеки или NewsEdge Service компании Dialog, ориентированный на пользователей, желающих оперативно получать новости по индивидуально сформированным запросам.

В перспективности цифровых библиотек можно не сомневаться. Быстрый доступ к достоверной и максимально полной информации - неотъемлемая составляющая фундаментальной науки, качественного образования и эффективного бизнеса. Ряд библиотек западных стран, прежде всего, библиотеки американских университетов, уже сегодня обеспечивают своим читателям возможность пользования подобными ресурсами. Постепенно это становится стандартной библиотечной услугой. Несомненно, что в будущем, с дальнейшим расширением в библиотеках доступа к

Интернет, подписка на полнотекстовые коммерческие базы данных будет являться составной частью бюджета, затрачиваемого на комплектование фонда.