

КЭРОЛ ТЕНОПИР
(ctenopir@utk.edu) – профессор Отделения библиотековедения и информации Университета штата Теннесси (г. Кноксвилл, США), постоянный автор рубрики «Online databases» журнала «Library Journal». Публикация перевода статьи К.Тенопир «Time To Redefine Database» осуществлена с любезного разрешения автора.

Перевод **Жанны Рысевой**



Что такое «база данных»?

Еще несколько лет назад мы с легкостью давали определение термина «база данных». В нашем понимании это был «постоянно обновляемый компьютерный файл, содержащий информацию, рефераты статей или тематические справки, систематизированные для свободного и быстрого использования» (ODLIS: Интерактивный словарь по электронным библиотекам и информации, www.wcsu.ctstateu.edu/library/odlis.html). Издатель базы данных (к примеру, рефератов по психологии или технической информации) нес ответственность за информационное наполнение (и, возможно, распространение печатных указателей), а фирма-поставщик (например, Dialog или SilverPlatter) трансформировала ее для облегчения поиска и обеспечения доступа.

Эта схема была приемлемой для нас начиная с 1970-х годов. В некоторых случаях крупные производители баз данных (например, H.W.Wilson или Information Access Company/Gale Group) выступают как собиратели и поставщики своей собственной информации, а также как составители информации других фирм, но они всегда придерживаются общепринятой схемы. Когда SilverPlatter, OCLC или Dialog утверждают, что они предлагают сотни баз данных, мы знаем, что на деле это означает представление посреднических услуг по широкому отбору различных информационных продуктов, поставляемых фирмами, ответственными за информационное наполнение.

Согласно Указателю Гейла по онлайн-овым, офлайн-овым и межсетевым базам данных (2001 г.), базы данных подразделяются на библиографические (фактографические/реферативные), справочные, статистические, полнотекстовые или других типов. Независимо от типа, библиотекари называют каждый из этих отдельных файлов базой данных.

Исключения из этой модели имелись всегда. Некоторые фирмы (например, Lexis-Nexis) изначально предпочитали более гибкую модель. Большинство информационных продуктов фирмы Lexis-Nexis идут непосредственно от первичного издателя, а реже – от вторичной компании, отвечающей за систематизацию содержания. Поэтому продукцию Lexis-Nexis в большей степени характеризует перечень включенных полнотекстовых документов, чем количество баз данных. Компания предпочитает термин «файлы» чаще, чем «базы данных», и использует термин «библиотеки» для обозначения больших групп файлов. При этом файлом может быть, как одно наименование журнала, так и несколько предметных или хронологических информационных источников.

■ «Трудный термин»

Уже долгое время термин «база данных» вызывал у

меня несогласие по ряду причин. Во-первых, только информационные специалисты пользуются им для обозначения файла, состоящего из отдельных сегментов информации, и даже мы используем его в совершенно другом значении, когда описываем программное обеспечение управления системой базы данных. Различие, проводимое иногда (особенно в Европе) между отдельными базами и банками (системами) данных, на самом деле едва уловимо. Более того, этот термин используется, чтобы описать огромное многообразие типов информационных источников. Указатель совершенно отличается от энциклопедии, однако их цифровые версии называются в том или ином случае базами данных. Термин «полнотекстовые базы данных» настолько широкий, что он теряет смысл. Полный текст включает в себя собрания статей из журналов и газет, а также содержания альманахов и даже книг.

Наконец, подсчет баз данных при сравнении систем никогда не был легкой задачей, поскольку каждая система могла бы структурировать то же самое содержание по-другому. Например, фирма Lexis-Nexis сортирует базу данных Medline в соответствии с постоянной характеристикой статей, например, статьи, которые освещают только гуманитарные исследования или написаны на английском языке, в то время как фирма Dialog структурирует эту же огромную базу данных хронологически.

■ Мегабазы данных

Возможно, пришло время пересмотреть термин «база данных», так как мы вступили в эру, в которой господствуют полнотекстовые статьи, связанные с библиографическими записями, гибридные веб-сайты, которые содержат все типы информации на одном сайте, гибко меняя содержание, в зависимости от подписного периода или потребностей отдельного пользователя. Оцифрованные указатели, журналы, энциклопедии или газеты предназначены для широкого круга пользователей. Даже границы между базами данных внутри отдельной системы – нечеткие.

Например, в четвертом квартале 2000 г. Gale Group объявила, что объединенная база данных InfoTrac будет предлагаться, как единая мегабаза данных под названием OneFile. Она обеспечивает доступ к библиографическим записям из более чем 6.200 журналов и газет и других электронных новостных средств массовой информации (СМИ), некоторые из них имеют файлы двадцатилетней давности. Около 2.700 периодических изданий также включают полнотекстовые статьи.

Есть сведения, что OneFile включает более чем 16 миллионов записей (или библиографических, или полнотекстовых) с доступным интерфейсом и поисковым инструмен-

том InfoTrac. Библиотеки, которые сейчас выписывают одну или более баз данных InfoTrac (например, Academic ASAP, LegalTrac или National Newspaper Index), знают, что базы данных объединяют в себе научные журналы и журналы широкого профиля, специализированную литературу и новостные источники информации.

■ Как избежать дублирования

Объем записей, предлагаемых в каждой из отдельных баз данных InfoTrac, мог бы оказаться значительно больше 16 миллионов, потому что, хотя InfoTrac и предлагает более 40 отдельных баз данных, в них много дублирующей информации. База данных «Academic ASAP», например, включает источники, которые входят в базу данных расширенной «Academic ASAP». База данных InfoTrac «Издания для детей» (Kids Edition) включает подсистему «Издания для студентов» (Student Edition). OneFile устраняет значительное дублирование.

OneFile использует множество предметных рубрик для сочетания разнообразной информации из отдельных баз данных и объединяет разные индивидуальные базы данных. Почти одна треть заглавий – это журналы широкого профиля, четверть составляют университетские журналы и еще одна четверть – это публикации по бизнесу и технике. Остальные заглавия – журналы по праву, периодические издания по здравоохранению, газеты или новостные СМИ.

Крупные библиотеки или консорциумы могут заинтересоваться подпиской на OneFile, но следует иметь в виду, что его стоимость велика, правда, она может быть меньше для библиотек, которые подписывают большинство баз данных, включенных в InfoTrac. Отдельные библиотеки могут по-прежнему осуществить подписку на специализированные базы данных InfoTrac.

По утверждению менеджера по продукции InfoTrac, библиотекам предпочтительнее подписываться на базу данных OneFile, чем на отдельные базы данных InfoTrac, так как в ней упрощен доступ к некоторым междисциплинарным темам и нет необходимости вести перекрестный поиск.

Мегафайлы – это более удобный термин для обозначения того, как библиотечные пользователи воспринимают интерактивные ресурсы. Пользователи рассматривают Интернет как огромный информационный ресурс. Библиотеки заинтересованы в привлечении внимания пользователей и зачастую им приходится напоминать своему руководству, что подписка на базы данных – это наиболее эффективный способ для обслуживания научного сообщества. Если еще 10 лет назад 16 миллионов статей в базе данных OneFile, могли бы ошеломить пользователей, то сейчас они научились рациональному отношению к такому объему информации.

Следует иметь в виду, что OneFile – это только часть разрастающихся, гигантских, многопрофильных баз данных, подобных тем, которые продают фирмы Lexis-Nexis и ProQuest – два главных конкурента фирмы Gale Group на рынке университетских библиотек. Итак, являются ли отдельные базы данных устаревшими? Специалисты утверждают, что традиционная модель издатель/распространитель трансформировалась, и сейчас на рынке подписных услуг отдается предпочтение подписке на онлайн-обслуживание.

■ Гибридные ресурсы

Web-схема и системная память способствовали осуществлению другого направления: объединению ряда различных баз данных в единый политематический ресурсный центр. Так, Health and Wellness Resource Center – самая последняя гибридная база данных Gale Group – интегрирует указатель и полные тексты из 400 медицинских журналов и 2.100 периодических изданий широкого профиля вместе с 800 брошюрами и множеством справочников по здравоохранению.

База данных Health and Wellness Resource Center при-

дет на смену базам данных Health Reference Center и включает в себя другие, появившиеся недавно ресурсные центры, включая «Бизнес и компании», «Историю», «Литературу», а также различные ресурсные центры для студентов. Все они объединяют указатели периодики, полнотекстовые базы данных, созданные Information Access Company/InfoTrac, со множеством справочников, изданных Gale Group. Это помогает пользователю найти основную информацию из энциклопедии или каталога, а затем более детальную информацию из периодических статей.

■ Двойной путь ProQuest

ProQuest (прежнее название – UMI/Bell & Howell Information and Learning) также продает информационное содержание в мегафайле или как отдельные базы данных. База данных ProQuest 5000 объединяет наиболее популярные библиографические и полнотекстовые записи из 5000 газет и журналов широкого профиля и более специализированных источников информации (к примеру, коммерческой базы данных ABI/INFORM).

Однако большинство абонентов ProQuest все еще выбирает индивидуальные базы данных. Специалисты ProQuest сообщают, что обычно библиотеки отдают предпочтение покупке больших интегрированных баз данных. Высоко оценивая такие отдельные базы, как ABI/INFORM, они затем приобретают практически все базы данных, которые предлагает фирма.

Это означает, что концепция отдельных баз данных пока еще не устарела. Библиотеки все еще используют отраслевую основу для оценки содержания. Не имеется никого готового решения на все случаи. Библиотеки по бизнесу нуждаются в основательных информационных продуктах по бизнесу, технические библиотеки – в научных базах данных по технике, правовые библиотеки – в базах данных по праву. Мегабаза данных подразумевает, что в библиотеках имеется место для огромной универсальной базы данных. Имея репертуар основных информационных продуктов, мы не можем быть уверены, что он отражает все области знания.

Специалисты считают, что больше – не всегда лучше. Они не расширяют содержание баз ProQuest только для увеличения ее объема. Базы данных ProQuest снабжены предметным указателем и структурированы для основных типов пользователей. В ABI/INFORM, например, указатели предназначены для пользователей, интересующихся вопросами бизнеса.

■ Новые информационные продукты H.W.Wilson

Компания H.W.Wilson также начала группировать свои значительные информационные ресурсы в мегабазы данных, из которых получают базы данных различного содержания. База данных Wilson OmniFile Full Text Mega включает все указатели (начиная с 1982), рефераты (начиная с 1984) и полные тексты (начиная с 1994) из шести баз данных H.W.Wilson (Education, General Science, Humanities, Readers' Guide, Social Sciences и Business Full Text) плюс полнотекстовые статьи из дополнительных пяти баз данных. База данных OmniFile Full Text Select включает полнотекстовые статьи, отобранные из множества полнотекстовых баз H.W.Wilson, только начиная с 1994 г. Дотошные редакторы H.W.Wilson снивелировали разночтения указателей среди баз данных OmniFile и улучшили поиск. Название компании H.W.Wilson и ее авторитетные файлы также обозначены в OmniFiles.

Даже совместимая индексация не позволяет утверждать, что мегабазы данных отвечают запросам пользователей. Web-сайт H.W.Wilson объясняет, что подход «поиск с одной остановкой» делает OmniFile особенно удобным для пользователей с запросами из перекрестных областей знаний, например, таких, как этнические исследования, женские исследования, история. Хотя они могут быть хороши и для пользователей, которые не знают, с чего начать, т.е. в