

Лапо П.М. (Минск)

**ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
опыт Иллинойского университета**

(Урбана-Шампейн, США)

Благодаря широкому внедрению компьютеров, развитию сетевых технологий и Интернета тема создания электронных библиотек является сегодня чрезвычайно актуальной, и не случайно, что доклады нашей секции в том или ином ракурсе ее освещают. С этой темой связаны и выступления, посвященные проблемам разработки национального MARC-формата для обмена библиографической информацией, и выступления по проблемам создания сводного электронного каталога (СЭК)

библиотек Беларуси и его интеграции с СЭЖ других стран. Поскольку первенство в создании обсуждаемых здесь библиотечных технологий принадлежит библиотекам США, то, на мой взгляд, будет небезынтересной и полезной информация об опыте наших американских коллег в деле развития электронных библиотек страны на современном этапе. Конечно, невозможно в одном докладе достаточно подробно охватить всю проблематику такого сложного явления, как “электронные библиотеки”. В меру своих сил мне хотелось бы осветить исследования, проводимые американскими специалистами в рамках национальной программы “Инициатива в области цифровых (электронных) библиотек” (“Digital Libraries Initiative (DLI)”), пользуясь знаниями и информацией, полученными мной во время полугодового обучения по профессиональной библиотечной программе при университете штата Иллинойс в г.Урбана-Шампейн (США) и стажировки при входящем в состав библиотечной системы университета информационно-библиотечном техническом центре им. Грейнджера, специалисты которого непосредственно участвуют в программе DLI.

Работы, выполняемые в рамках данной программы, видятся мне как перспективы развития библиотечного дела и в нашей стране, и хотелось бы уже сегодня привлечь к ним внимание белорусских специалистов в области библиотечного дела, чтобы наши решения были более продуманными, обоснованными и учитывали опыт других стран.

Программа “Digital Libraries Initiative”

Национальная исследовательская программа США “Digital Libraries Initiative (DLI)” была начата в сентябре 1994 г. Она финансируется Национальным научным фондом США (NSF), Агентством по поддержке проектов перспективных исследований (ARPA) и Национальной администрацией по авиации и космосу (NASA) — в среднем 1 миллион долларов США в год — и должна быть выполнена к сентябрю 1998 г. Следует отметить, что данная программа является основной исследовательской частью такой важной для федерального правительства и всей нации программы, как создание Национальной информационной инфраструктуры (National Information Infrastructure (NII)), конечная цель которой — *обеспечить каждому американцу доступ к необходимой ему информации, где бы он ни находилась* [1, с. 22].

Таким образом, библиотеки в США как рассматривались, так и рассматриваются в качестве основных элементов си-

системы информационного обслуживания как различных структурных элементов американского общества, так и широких слоев населения страны. Современные информационные технологии, доступность компьютеров и развитие сетевых инфраструктур не отменяют ни одной из присущих библиотеке и выполняемых ею на протяжении столетий функций, но позволяют поднять информационное обслуживание на качественно новый уровень, чем и вызвано появление такого термина, как “электронная библиотека” или более употребительного в США — “цифровая (*digital*) библиотека”. Существует довольно большое число его определений. Применительно к проекту DLI “электронная библиотека” — это библиотека, которая хранит информацию на электронных носителях и эффективно манипулирует большими собраниями такой информации, причем сами собрания могут быть распределены по сети Интернет. И в то время, как реальные электронные библиотеки должны сконцентрировать свое внимание на вопросах стоимости доступа к источникам информации и технологиях перевода информации в электронную форму хранения, исследования, выполняемые в рамках DLI, ставят во главу угла вопросы, как развить необходимую информационную структуру и какие нужны механизмы для эффективного поиска информации в сети.

Рабочая группа по технологиям информационной инфраструктуры и ее приложениям (Information Infrastructure Technology and Applications (IITA) Working Group), представляющая собой высший уровень технического комитета по созданию НИИ, провела в мае 1995 г. семинар, на котором были определены основные направления исследований в области создания электронных библиотек, которые видятся как сеть распределенных репозитариев, представляющих собой индексированные коллекции, в которых объекты любого типа могут быть найдены, несмотря на различия существующих сетевых протоколов и форматов хранения [1, с. 23]. Была поставлена цель развития технологий от — в ближайшем будущем — “прозрачности” поиска, несмотря на различные форматы и протоколы, до — в более отдаленной перспективе — “прозрачности” поиска информации, несмотря на различия в ее описании и толковании. Как задача-максимум было определено достижение в рамках создания электронных библиотек “глубокого семантического взаимодействия (*interoperability*), т.е. обеспечения пользователя возможностью “ясного и логически последовательного поиска среди автономно определенных и поддерживаемых классов объектов и служб, распределенных

в виде разнородных репозитариев, на основе использования программного обеспечения, которое бы объединило такие репозитории или выступило посредником при доступе к ним, несмотря на какие-либо присущие репозитариям различия... Достижение этого потребует прорывов в описании информации так же, как в разработке протоколов информационного поиска, обмена объектами и поиска объектов. Проблематика здесь включает определение и использование метаданных об объектах и их сбор или вычисление на основе самих объектов (представленных как в виде текста, так и в форме мультимедиа); объединение (*federation*) и интеграцию разнородных репозитариев с несравнимыми семантиками; группировку и автоматическую организацию иерархической структуры информации; а также алгоритмы для автоматической классификации, ранжирования и оценки качества информации, ее жанра и других свойств" [1, с. 24].

Столь длинной цитатой мне хотелось бы дать моим коллегам-библиотекарям представление о том поле деятельности, которое открывается в современных условиях существования и развития библиотек и, если можно так выразиться, подготовить морально к тем испытаниям, которые ждут нас уже в ближайшем будущем. Понятно, что для того, чтобы библиотеки и дальше занимали достойное место в информационном мире, они должны овладеть приведенным в цитате понятийным аппаратом, чтобы использовать его в своих исследованиях и исходить из него в своей практической деятельности.

И еще на одну особенность выполнения проектов в области создания и развития электронных библиотек обратила внимание их разработчиков вышеупомянутая Рабочая группа. Это необходимость учета большого масштаба как самих репозитариев, так и связывающей их сети, большой разброс типов объектов и предметных областей, которым они принадлежат. Вот почему все реализуемые в рамках DLI проекты выполняются на крупномасштабных коллекциях информации (так называемых *testbed collections*).

Проекты, выполняемые в рамках DLI

В результате конкурса из 73-х заявленных на участие в программе DLI проектов было выбрано 6. Это проекты:

- университета Карнеги-Меллона;
- Калифорнийского университета в Беркли;
- Калифорнийского университета в Санта-Барбаре;
- Иллинойского университета в Урбане-Шампейне;
- Мичиганского университета;

— Стенфордского университета.

Все шесть участников используют разные подходы к решению проблем, связанных с созданием и развитием электронных библиотек, и ставят перед собой в своих исследованиях разные задачи.

Так, проекты Иллинойса и Беркли предполагают создание законченных информационных систем со многими пользователями: Иллинойсский проект сфокусирован на созданных вручную структурированных текстовых документах, а проект Беркли — на автоматически распознаваемых образах документов. Эти проекты используют дополняющие друг друга подходы: в первом случае получают материалы в электронном формате в виде SGML документов и графических файлов в PDF-формате напрямую от издателей; во втором — получают от издателей печатные материалы в больших количествах и автоматически трансформируют статьи журналов в электронную форму.

Университеты Карнеги-Меллона и Санта-Барбары планируют предоставить новые средства поиска и хранения информации, которая имеет форму, ранее не позволявшую ее индексировать и искать. Университет Карнеги-Меллона исследует проблемы сегментации и индексирования видео, используя автоматическое распознавание речи и знание структуры записанной информации. Университет Санта-Барбары индексирует географические карты, используя автоматическую обработку образов и информацию о *метаданных* региона.

Наконец, проекты Стенфордского и Мичиганского университетов исследуют проблемы так называемых *gateways*, стыковочных узлов разнородных компьютерных сетей, которые необходимы для выполнения поиска информации в крупномасштабных электронных библиотеках. Выполнение обоих проектов направлено на разработку механизмов, которые, с одной стороны, позволяли бы взаимодействовать различным поисковым службам независимо от форматов хранения и передачи данных, а с другой — находить соответствующие запросу источники информации.

Все проекты имеют крупномасштабные коллекции для проведения фундаментальных исследований в области построения электронных библиотек в сети.

Среди всех аспектов исследования проблем создания и развития электронных библиотек разработчики особо выделяют индексирование (*indexing*) и объединение (*federating*), поскольку поиск информации в электронной библиотеке — это поиск в проиндексированной коллекции объектов, зачастую

распределенной в сети. Таким образом, *репозитарий* — это проиндексированная коллекция объектов. Процесс же их объединения — это установка связей (*mapping*) между схожими по каким-либо признакам и хранящимися в разных репозитариях объектами таким образом, что конечным пользователем они воспринимаются как принадлежащие к одной коллекции-репозитарию.

И процесс индексирования, и процесс объединения репозитариев упираются в проблемы семантики. Брюс Шац (Bruce Schatz), один из ведущих специалистов Иллинойского университета в области проблем информационного сетевого поиска, в одном из своих выступлений отметил, что в тенденции развития информационного поиска в сети он бы выделил следующие этапы:

- для 1970—90-х гг. был характерен поиск текста, приоритетное внимание уделялось синтаксису представления информации;

- для 1980—90-х гг. характерен поиск документов и переключение внимания на структуру данных;

- для конца 1990-х и до 2010 г. будет характерен концептуальный поиск и обращение внимания на семантику данных.

Технология информационного поиска не изменилась принципиально за последние 30 лет и до сих пор служит миллионам пользователей сети Web. Как правило, она обращает внимание на синтаксис информации. Обычно запрос с языка пользователя конвертируется в язык запроса каждого поискового механизма для каждого индекса коллекции, в лучшем случае, учитывая структуру метаданных, отражающих информацию рассматриваемой коллекции. То есть информация в поле “автор” (AUTHOR) может ставиться в соответствие с различными полями, например, такими, как AU или AUT для различных коллекций данных. Однако непредусмотренные значения могут быть проигнорированы таким конвертором.

Следующий этап в развитии информационного поиска обещает семантический анализ информации и создание систем, позволяющих находить связанные по смыслу части в большом числе распределенных источников информации. Такой подход в какой-то мере реализуется на базе использования стандарта SGML применительно к текстовой информации, т.е. только части информационных ресурсов, хотя и наиболее изученной.

Проекты DLI призваны на основе широкомасштабных коллекций данных и партнерских связей создать новые технологии поиска информации, где бы она ни хранилась, с уче-

том всей ее многогранности. "Сегодня на Web вы ищете информацию, просматривая документы. Завтра на Web вы найдете ее, просматривая репозитарии. В новом тысячелетии, помимо Web, вы будете иметь технологию анализа окружающей среды, которая позволит вам скоррелировать информацию из разных репозитариев для решения ваших проблем" [1, с. 26].

Конечно, приведенная здесь краткая, можно сказать, беглая характеристика разрабатываемых в рамках DLI проектов не отражает все многообразие проблем, с которыми сталкиваются их разработчики. Полученный по мере реализации проектов опыт также трудно вместить в узкие рамки доклада. Поэтому и прилагается список существующих в Интернете источников информации о DLI — проектах.

Проект DLI Иллинойского университета

В настоящее время основным источником информации все еще остается текстовый документ. Проект DLI Иллинойского университета направлен на развитие информационной инфраструктуры для эффективного поиска научно-технической литературы в Интернете. Экспериментальный репозитарий (*testbed collection*), на котором обрабатываются технологии поиска, содержит свыше 4-х тыс. статей из журналов, посвященных проблемам компьютерной техники, электроники, физики, гражданского строительства, аэрокосмических исследований. Издатели поставляют статьи журналов в SGML (*Standard Generalized Markup Language*) и PDF (*Portable Data Format*)-форматах (для подготовки статей в формате SGML используются средства, разработанные фирмой ArborText), специалисты университета индексируют их и обеспечивают интерактивный доступ к ним на основе использования Web интерфейса. SGML формат рассматривается как перспективный язык открытых систем документов. PDF формат используется для передачи и представления встречающейся в статьях графической информации (графиков, схем, рисунков и т.п.). Основным инструментальным средством, используемым для разработки программного обеспечения системы хранения и поиска информации, является Visual Basic. Но вполне возможно, что он будет вытеснен языком Java, который по мере своего развития становится все более и более привлекательным для решения такого рода задач. Для просмотра подготовленных издателями в SGML формате статей используется Panorama SGML viewer, разработанной компанией SoftQuad.

Проект DLI выполняется, прежде всего, специалистами информационно-библиотечного технического центра им. Грейндже-

ра, но, вместе с тем, в него вовлечены и специалисты других подразделений университета, например, Национального центра суперкомпьютерных приложений, библиотечного факультета университета и др. В частности, преподаватели и студенты библиотечного факультета выполняют часть проекта, связанную с учетом мнения пользователей разрабатываемых университетом в рамках проекта DLI информационных технологий. Это тоже многоаспектное исследование, целью которого является изучение того, как научно-техническая деятельность и учеба связаны с использованием распределенной информации в электронной форме. Апробируются различные экономические модели предоставления доступа к информации, выполняется индивидуальное интервьюирование пользователей, ведется работа с группой пользователей по обсуждению наиболее важных проблем, существующих вокруг доступа к информации с научной или образовательной целью. Такие социологические разработки привели уже к определенным результатам. Так, например, многие профессора отметили, что по диаграммам, приведенным в статье, они лучше, чем после чтения резюме статьи или заключения, могут определить, представляет для них данная статья интерес или нет. По их мнению, диаграммы и расчеты, приведенные в статье, лучше раскрывают то, что реально сделано ее автором, а не его желаемые достижения. Несколько опрошенных заявили о том, что иногда уравнения, приведенные в статье, были той единственной ее частью, в которой они нуждались для выполнения своей работы. Эти предварительные результаты исследований подтверждают мысль, что эффективность работы с электронной библиотекой зависит от того, насколько гибким будет интерфейс поиска информации и обработки структуры документа [2, с. 32].

Выводы

Суммируя все вышесказанное, хотелось бы акцентировать внимание на следующем:

— проекты создания электронных библиотек финансируются в весьма значительных размерах и выполняются в рамках национальной программы DLI, которая, в свою очередь, является основной составной частью общенациональной программы по созданию Национальной информационной инфраструктуры — основы информационного общества США;

— проекты DLI адресованы решению проблем, с которыми белорусским библиотекам предстоит столкнуться в ближайшем будущем. Хотелось бы, чтобы опыт американских коллег, уже не один год ведущих исследования в данном на-

правлении, был уже в ближайшее время проанализирован и при необходимости востребован;

— проекты DLI направлены на достижение глубокого семантического взаимодействия больших разнородных баз данных, расположенных в региональных и глобальных компьютерных сетях. Это очень сложная проблема, и потому необходимо уже сейчас, пока наши библиотеки еще только в начале создания разделяемых информационных ресурсов, приступить к ее решению, тщательно продумав технологию и базу стандартов для “прозрачной” стыковки информационно-библиотечных ресурсов Беларуси с международными;

— объектом исследований в проектах DLI являются не только технологии создания, хранения, поиска и выдачи результатов поиска нужной информации, но и люди, конечные пользователи этих технологий. Таким образом, наши библиотечные системы должны быть человеко-ориентированными, учитывать специфику, особенности информационных запросов различных групп пользователей, а в идеале — каждого пользователя.

И последнее. В выполнение таких высокотехнологичных проектов развития информационного обслуживания вовлечены сами библиотеки. Не раз с разработчиками проектов DLI обсуждался вопрос о том, отменяет ли реализация электронных библиотек библиотеки традиционного типа, и каждый раз в ответ звучало: “Нет, не отменяет”. Наоборот, она предполагает расширение функций и возрастание роли последних, что, в свою очередь, конечно, невозможно без овладения библиотечными работниками всем арсеналом технических и программных средств, применяемых в современных информационных технологиях.

ЛИТЕРАТУРА

1. Bruce Schatz, Hsinchun Chen. Building Large-Scale Digital Libraries. — Computer, May 1996, Vol. 29, No. 5, p. 22—26.

2. Bruce Schatz, William H. Mischo, Timothy W. Cole, Joseph B. Hardin, Ann P. Bishop, Hsinchun Chen. Federating Diverse Collections of Scientific Literature. — Computer, May 1996, Vol. 29, No. 5, p. 28—36.

Информация об участниках проектов DLI и самих проектах в Интернете.

3. Университет Карнеги-Меллона: <http://fuzine.mt.cs.cmu.edu/im/informedia.html>.

4. Стенфордский университет: <http://Walrus.Stanford.EDU/diglib/>.

5. Калифорнийский университет в Беркли: <http://elib.cs.berkeley.edu/>.

6. Калифорнийский университет в Санта-Барбаре: <http://alexandria.sdc.ucsb.edu/>.

7. Иллинойский университет в Урбане-Шампейне: <http://www.grainger.uiuc.edu/dli>.

8. Мичиганский университет: <http://http2.sils.umich.edu/UMDL/HomePage.html>.