

Вопросы организации электронной библиотеки на основе корпоративной компьютерной сети

Е.В.Ковязина

Институт вычислительного моделирования СО РАН

Создание корпоративной компьютерной сети Красноярского научного центра (КНЦ) СО РАН подготовило базовые условия для организации на её основе электронной библиотеки, включающей ресурсы всей сети библиотек институтов. Основная идея, которой придерживались её организаторы, была «получение всей необходимой информации прямо на рабочем месте, без посещения библиотеки». В сущности, это означало бы построение распределённой электронной библиотеки, содержащей все имеющиеся в библиотеках КНЦ информационные ресурсы в их полном виде. Несмотря на очевидную невозможность полномасштабной реализации этой идеи, в силу неразрешимых юридических, экономических и т.п. проблем, создание технологической и технической основы для такой библиотеки представляется вполне возможным, учитывая современное состояние развития этой области. В результате должна получиться гибридная библиотека, сочетающая черты как электронной, так и традиционной библиотеки.

Под технической и технологической основой (ТТО) здесь понимается совокупность программных и технических средств, позволяющая реализовать любую из компонент, составляющих электронную библиотеку. Можно выделить следующие группы ТТО:

1. Единый интерфейс доступа к распределённым информационным ресурсам с возможностью поиска по широкому набору параметров
2. Стандартизированные средства описания разнородных информационных ресурсов с обязательными инструментами доступа к их полному электронному образу
3. Средства оцифровки документов и преобразования графических файлов в один из стандартно используемых в Интернет форматов
4. Средства защиты информационных ресурсов и контроля их использования
5. Средства взаимодействия библиотек, обеспечивающие возможность обмена, как документами, так и отдельными записями баз данных, и обеспечивающие работу с распределёнными ресурсами
6. Средства автоматизации технологических процессов внутри библиотеки

Для решения этой задачи на первом этапе группы ТТО были соотнесены с основными компонентами электронной библиотеки, с целью определить их способность реализовать каждую из них. Разделение электронной библиотеки на компоненты, описание их, вместе с описанием стандартов, используемых в этой области, наиболее полно дано в докладе Р.Меррея [1] на конференции «Онлайн-99». Были исследованы потребности библиотек и их читателей по каждой из этих компонент. Исследованы отечественные и зарубежные программные разработки, их функциональное наполнение и возможности в области их предполагаемого использования, а также и условия их получения. Для соответствия создаваемой системы российским и мировым стандартам, а также и потому что в Красноярске не существует исследовательских групп в области автоматизированных библиотечных технологий, было принято решение

использовать промышленные программы и принятые в России стандарты, дополняя их собственными разработками только в случае крайней необходимости.

Для достижения поставленных целей в разрабатываемой системе используется ИРБИС (ГПНТБ России), WWWSearch (ГПНТБ СО РАН) и ZooPARK (ИГТиМ СО РАН). В совокупности с возможностями ОС Windows NT Server - Workstation, MSOffice, программами распознавания отсканированного текста и набором средств подготовки и чтения текстов различных, принятых в Интернет, форматов, они позволяют решить поставленную задачу. Построенная на основе этих программных продуктов система обладает значительной функциональной избыточностью, то есть одну и ту же операцию можно произвести несколькими различными программными средствами. Причём система имеет тенденцию к увеличению избыточности, и как следствие её чрезмерному усложнению, по мере дальнейшего совершенствования отдельных программных частей. Кроме трудностей, связанных с обучением персонала библиотек, это влечёт за собой значительные трудозатраты по обслуживанию системы, как единого целого, требующие привлечения специалистов, которых библиотеки, как правило, не имеют.

Надо признать, что любая библиотека, включаясь в процесс автоматизации, и приобретая программное обеспечение, ожидает, что оно способно решить все её проблемы при минимальном привлечении специалистов «со стороны». Хотя при высокой квалификации персонала и должном его обучении это действительно так, но есть в этой области задачи, которые, с одной стороны, являются локальными задачами каждой библиотеки, отражая её специализацию, с другой стороны, требуют достаточно высокой специальной квалификации для их решения. По мере развития системы проявляются всё новые и новые задачи такого рода, требующие решения своими местными силами, и окончания этого процесса не предвидится. Наиболее трудоёмкие из них, которыми уже длительное время занимается библиотека Института вычислительного моделирования СО РАН, можно условно разбить на две группы:

1. Задачи, связанные с функциями компьютерной сети.
2. Задачи, связанные с лингвистическим обеспечением.

К первой группе задач относятся задачи контроля использования информационных ресурсов и защиты информации в сети. Возникновение этих задач связано со спецификой библиотеки академического института. Как правило, каждая библиотека, входящая в сеть библиотек академического филиала, - это библиотека с небольшим специализированным фондом, занимающая ограниченное по площади помещение и имеющая маленький штат. Библиотека имеет компьютеры, обеспечивающие работу библиотечного web-сервера, и рабочие места персонала. В подавляющем большинстве библиотек КНЦ СО РАН нет ни одного читательского рабочего места. Более того, задача создания таких рабочих мест даже не ставится. В качестве читательских рабочих мест, а зачастую и рабочих мест каталогизаторов, используются все компьютеры, входящие в корпоративную сеть КНЦ СО РАН и локализованные внутри соответствующего Института. Это избавляет библиотеки от многих проблем, но требует обеспечения контроля использования информационных ресурсов того же уровня, который имеет место в компьютерном читальном зале внутри библиотеки, но с большей степенью автоматизации, так как отсутствует визуальный и личный контроль со стороны персонала библиотеки. Это группа

задач может быть решена только при непосредственном активном участии, как администратора корпоративной сети, так и администраторов сетей каждого из институтов, входящих в корпоративную сеть, так как её решение требует установки специальных средств на проху-серверах и редактирования администраторских баз данных. Защита собственных информационных ресурсов реализована внутри Институтов средствами операционных систем на библиотечных серверах, определением прав доступа пользователей. Контроль доступа к информационным ресурсам библиотек извне по НТТР, набор баз данных доступных внешним пользователям – все эти задачи вполне успешно решает администраторская часть системы WWWSearch, она же и ведёт полную статистику обращений к информационным ресурсам библиотеки. Остаётся открытой задача контроля обращения читателей библиотек к внешним базам данных, таким как полнотекстовые журналы, доступные библиотекам как членам консорциума библиотек и по лицензионным соглашениям, содержащим те или иные ограничения на использование этих ресурсов. Для того чтобы предоставить возможность пользователям свободно обращаться к необходимым им информационным источникам, в качестве IP-адреса для доступа указывается адрес проху-сервера корпоративной сети, посредством которого подавляющее большинство компьютеров в сети получает выход в Интернет. Для того чтобы исключить злоупотребления со стороны пользователей на проху-сервере организуется контроль обращений по перечню адресов, определённому совместно библиотеками КНЦ СО РАН.

В процессе каталогизации создаётся подробное описание каждого документа, независимо от того в каком виде он существует в библиотеке. Описание документа непременно должно соответствовать стандарту и содержать все максимально возможные поля. Описание документа может быть создано, как непосредственно в библиотеке, так и заимствовано в другой вышестоящей библиотеке или центре каталогизации. В случае заимствования запись должна дополняться собственной информацией библиотеки, получившей её. Следует заметить, что в условиях, когда поиск информации по запросам пользователя осуществляется в электронном каталоге, каждая запись есть не только простое описание документа, а и его поисковый образ. И чем правильнее он сформирован, тем эффективнее будет поиск информации в дальнейшем. С созданием поискового образа документа и связана вторая группа задач, требующих решения. Поисковый образ документа (ПОД) – это совокупность параметров, позволяющая отыскать необходимый документ по запросу пользователя. Обычно ПОД включает автора, заглавие, тип и год издания и т.д. Чем лучше проработан и стандартизован каждый параметр, который используется в поиске, тем эффективнее поиск документа. В том виде, в каком ПОД сформирован сейчас, его низкая эффективность при поиске информации очевидна. Поиск ведётся по автору, словам из заглавия, году, издательству и предметным рубрикам. Предметные рубрики берутся из ГРНТИ и отличаются крайне низкой детализацией, что особенно заметно при поиске в узкоспециальной области, который как раз и преобладает в академической библиотеке. Для того чтобы детализировать ПОД существует раздел «ключевые слова». Для заполнения этого раздела необходим нормализованный дескрипторный словарь, включающий лексику из областей знаний, соответствующих специализации библиотек. К сожалению, если такие словари и существуют где-либо, то они не продаются, не распространяются, и чаще всего являются словарями внутреннего пользования библиотеки, в которой они

созданы. Для решения этой проблемы в сети библиотек КНЦ была даже освоена методика создания и пополнения дескрипторного словаря. В результате чего стало ясно, что работа эта крайне трудоёмкая, а главное, не позволяет решить проблему стандартизации поиска документа. Ведь даже при наличии своего собственного словаря, при помощи которого будет создан адекватный поисковый образ читательского запроса, он не поможет при поиске в базах данных других библиотек. Если при создании поискового образа документа на бумажном носителе так остро стоит только проблема определения ключевых слов, то эта проблема многократно увеличивается при описании ресурсов на нетрадиционных носителях. При наличии рекомендаций не очень трудно создать рабочий лист, например, для описания ресурсов на электронных носителях. Но существуют поля, содержательная часть которых не определена, или неизвестна библиотекам, описывающим ресурс, по причине отсутствия таковой в методических рекомендациях. Значит невозможно подключить справочник-подсказку, что ведёт к произволу в описании, зависящему только от лица, описывающего ресурс. Ещё острее проблема при описании электронных карт и атласов, которые не могут быть созданы по аналогии с картами на бумажных носителях, ведь даже масштаб их не определён. Здесь справочный аппарат и стандарты на него вообще отсутствуют.

В процессе создания ТТО возникшие задачи не являются единственными трудностями, но это самые трудоёмкие, а порой и неразрешимые проблемы. Хотя, очевидно, не последние, требующие решения.

Список литературы

- [1] Меррей Р. Компоненты цифровой библиотеки и их взаимодействие// Научные и технические библиотеки. 2000. №6. С.56-68.