

водится наладка новой программной системы и перевод ЭУК в эту систему.

Создание лабораторного практикума ведется в среде графического программирования LabView, использование которой в вузе регламентируется лицензионным соглашением уровня «Университет» с компанией National Instruments. Данная среда позволяет за короткий промежуток времени смоделировать виртуальный лабораторный стенд и предоставить доступ к нему из ЭУК. Создаваемые виртуальные лабораторные работы подключаются к LearningSpace с применением стандарта SCORM, таким образом, обеспечивается автоматическая регистрация студента в лабораторной работе и затем сохранность результатов его работы в БД. Ведутся работы по созданию виртуальных работ с доступом к реальным лабораторным стендам посредством LabView непосредственно из системы дистанционного обучения.

Воройский Ф.С.

ТРЕБОВАНИЯ К АНАЛИТИЧЕСКОЙ ОБРАБОТКЕ ДОКУМЕНТОВ В ИНТЕРЕСАХ ОБЕСПЕЧЕНИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ И РАЗРАБОТОК

fsv@gptn.ru

Московский государственный университет культуры и искусств

г. Москва

Аналитическая обработка документов и формирование проблемно-ориентированных баз данных (ПОБД) в библиотеках и информационных органах может производиться для ИО и СИО конкретных коллективов ученых и разработчиков и/или для более широкого их использования через локальный или телекоммуникационный доступ, как в свободном, так и

платном режиме. В последнем случае всем видам пользователей должно быть предоставлено в свободном доступе достаточно точное и представительное описание ресурса для его нахождения и принятия решения о приобретении, а сам ресурс предоставляется на платной основе. При этом важнейшими показателями эффективности работы служб, производящих аналитическую обработку документов и ПОБД для включения их описаний в электронные каталоги автоматизированных систем, являются точность и полнота поиска.

«Традиционными» и наиболее используемыми в большинстве российских библиотек являются ИПЯ классификационного типа, предназначенные для классификации и предметизации документов. В информационных органах, работающих в сфере научно-технической информации, предпочтение исторически отдается вербальным (дескрипторным) языкам индексирования. Основные различия в этих ИПЯ и реализующих их словарных средствах (с одной стороны классификационных таблицах, с другой — тезаурусов и ключевых слов) заключаются в следующем:

- 1) Классификационные языки, являющиеся «предкоординатными», оперируют терминами, составляющими «рубрики» или «подрубрики», т.е. заголовками, описывающими некоторые области или подобласти, в соответствующей области знания. Дескрипторные языки, являющиеся «(пост)координатными», ориентированы на точное указание объекта(ов) описания или поиска, в том числе — на фактографические сведения;
- 2) Статьи в рубрикаторах, описывающих наименование рубрик и подрубрик любого уровня, могут составлять сложные словосочетания и предложения. Лексические единицы дескрипторных языков строятся на коротких словосочетаниях (1-3 слова). При необходимости построения составных словосочетаний (например, в поисковых пред-

писаниях) используются логические операторы, задающие разные виды отношений между лексическими единицами [1, 2, 3, 4, 5].

Таким образом, следует подчеркнуть, что вербальное индексирование позволяет описать разные аспекты содержания документов и БД с большей разрешающей способностью и более точно, чем классификационные ИПЯ. Исходя из изложенного, следует подчеркнуть важность использования для удовлетворения информационных потребностей ученых и разработчиков вербального (дескрипторного) индексирования, которое, кстати сказать, лежит в основе работы всех ИПС. При этом мы ни в коей мере не отвергаем необходимость классификационного индексирования, которое имеет важное значение в первую очередь для обеспечения различных внутрисистемных процессов (в библиотеке, группе библиотек, консорциуме и т.п.).

Традиционными процессами аналитической обработки документов и формирования локальных и распределенных электронных каталогов (ЭК) в библиотеках и информационных органах является составление библиографических записей (БЗ). Напомним, что в соответствии с **ГОСТ 7.1-2003** БЗ включает библиографическое описание («совокупность библиографических сведений о документе, приведенных по определенным правилам, устанавливающим наполнение и порядок следования областей и элементов, и предназначенных для идентификации и общей характеристики документа»), заголовок библиографической записи (в соответствии с **ГОСТ 7.1-2003**), термины индексирования, аннотацию или реферат (см. **ГОСТ 7.9-95**), шифры хранения документа, справки о добавочных библиографических записях, дату завершения обработки документа, сведения служебного характера. Реализация указанных требований предусмотрена действующими коммуникативными форматами серии MARC, предусматриваемыми в составе БЗ соответствующих полей данных, в частности в

UNIMARC`е и RUSMARC`е для предметных рубрик — 606 и индексов классификации — 675, 676, 679 и 686; для ключевых слов и персоналий — 610 и 600; для аннотаций и рефератов — 330 и т.п. [2, 6].

С большим сожалением следует отметить, что указанные требования выполняются в библиотеках не в полном объеме и далеко не всегда в удовлетворительном качестве:

- 1) Несмотря на то, что сотрудники подавляющего числа библиотек по данным анкетирования, проведенного среди участников АРБИКОНа в 2004 г. (более 150 библиотек разной ведомственной принадлежности) признают важность индексирования документов ключевыми словами (КС), реально используется этот способ аналитической обработки документов далеко не во всех библиотеках;
- 2) Качество индексирования КС по полноте и корректности отражения важных аспектов содержания документов также в большинстве случаев оставляет желать лучшего. В значительной степени это связано со следующими обстоятельствами: 1) Приверженностью сотрудников библиотек (особенно федеральных и вузовских) к традиционным способам и средствам индексирования (имеются ввиду «систематизация» и «предметизация») и в том числе — частая путаница принципов «координатного индексирования» с «предметизацией» (подробно об этом нами говорилось в докладах на международных конференциях КРЫМ 2004 и LIVCOM 2004 [7, 8]); 2) отсутствием согласованной методики составления КС (требования ГОСТ 7.66—92. «Индексирование документов. 3) Отсутствием общепринятой нормативной терминологической базы для вербального индексирования, в частности тезаурусов и «авторитетных записей». Общие требования к координатному индексированию» явно недостаточно)

— существуют разные методики, которые в ряде важных положений существенно противоречат друг другу [3, 7, 8];

- 3) Так же недостаточной является аналитическая обработка документов во многих библиотеках в части описания содержания документов для пользователей, поскольку в большинстве БЗ отсутствуют аннотации (не говоря уж о рефератах).

Очевидно поэтому, что гигантский объем литературы и содержащейся в ней информации остается не востребованным для пользователей как локальных, так и распределенных библиотечно-информационных систем. Особенно для тех из них, которые профессионально заняты в разных областях научных исследований и разработок.

Еще более серьезные требования к аналитической обработке документов и данных ставятся для их эффективного использования в режиме удаленного доступа. К ним относятся:

- 1) Требования обеспечения информационной совместимости в первую очередь для электронных каталогов и составляющих их БЗ это корректное использование коммуникативных (обменных) форматов и, в первую очередь — RUSMARC. А поскольку большинство внутренних форматов ПО АБИС ориентировано на форматы UNIMARC и UMARC 21 (USMARC), к тому же неоднозначно трактуемые, возникают проблемы совместного использования библиотечно-информационных ресурсов как библиотеками, так и их пользователями.
- 2) Описания отдельных полнотекстовых ресурсов, предоставляемых электронными библиотеками в Интернет или Интранет, созданными на базе библиотек и информационных органов, все более настоятельно требуют использования специальных систем метаданных и

коммуникативных форматов, которые отражают виды этих ресурсов, характер их принадлежности, использования и т.п.

Термин «**метаданные**» принято толковать, как «**данные о данных**», однако его значение распространяется помимо описания состава данных, их структуры (формата) представления, места хранения и других признаков описания также на поддерживающие их информационные системы, технологии, пользователей, методы доступа и т.д. Особенно широко этот термин стал использоваться в последние годы в связи с развитием **электронных библиотек**, поскольку метаданные стали важнейшим средством обеспечения навигации, поиска и возможности информационного обмена в Интернете. Однако до настоящего времени значение этого термина до конца четко не определено. Наиболее размыты границы между **метаданными** и **коммуникативными (обменными) форматами**.

Существуют различные категории метаданных, например, **описательные метаданные** (в том числе библиографические); **метаданные о структурах и форматах**; **административные метаданные**, содержащие данные для управления доступом; **идентификационные метаданные**, которые однозначно идентифицируют объекты внешнего мира и т.п. Помимо сказанного, метаданные подразделяются на те, которые предназначены для автоматического решения определенного класса задач — **машиночитаемые метаданные** и тех задач, которые решаются с участием человека — **человекочитаемые метаданные**.

Наибольшее развитие в мировой информационной практике и России получила система **Dublin Core, DC** — **Дублинское ядро (ДЯ)**, полное наименование системы: «Метаданные Дублинского ядра для простого открытия ресурса». Разработка ведется с 1995 г. рабочей группой с одноименным названием. Ею предложена простая структура описания документов, которая, по мнению разработчиков, должна заменить сложные

системы существующей каталогизации документов. Она предназначена для записи базовых структурных значений описания документов на языках разметки **HTML** и **XML**. Их состав включает в себя пятнадцать элементов, семантика которых была совместно определена международными группами профессионалов в области библиотечного дела, вычислительной техники, кодирования текстов, специалистов музейного дела и других смежных областей наук.

Действующие в Интернете и Рунете поисковые системы (Yandex, Rambler, Google, Yahoo!, Aport и др.) производят систематическое сканирование информационной среды Интернета, автоиндексирование и анализ выставленных ресурсов и обеспечивают пользователям нахождение нужных для них документов и данных. В указанном плане востребованность ресурсов и связанный с этим профессиональный (а может быть и коммерческий) успех организаций-создателей и владельцев этих ресурсов находятся в прямой зависимости от требований изложенных выше.

Кратко подводя итоги сказанного, следует подчеркнуть, что требованиями к аналитической обработке документов в интересах обеспечения научных исследований и разработок является детальное и корректное описание их содержания с использованием ключевых слов и выполнение требований коммуникативного формата RUSMARC. Средством технологической поддержки этого процесса должна стать согласованная инструкция по индексированию документов ключевыми словами.

Становится все более очевидным, что отдельным даже очень крупным библиотекам и информационным органам в одиночку не возможно удовлетворительно решить указанные задачи. С этой целью и создаются корпоративные объединения библиотек. В частности в рамках Ассоциации региональных библиотечных консорциумов (АРБИКОН) определены обязательные для всех участников «профили» для БЗ в формате RUSMARC,

ведутся работы по созданию единой методики координатного индексирования документов, рассматривается вопрос об участии совместно с национальными библиотеками в работе по созданию авторитетных записей, поставлен вопрос о начале рассмотрения вопросов, связанных с использованием метаданных.

Библиографический список

- 1) Жданова Г.С., Колобродова Е.С., Полушкин В.А., Черный А.И. Словарь терминов по информатике / под ред. докт. Техн. наук, проф. А.И. Михайлова. — М.: Наука, 1971. — 359 с.
- 2) ГОСТ 7.1.—2003. Межгосударственный стандарт. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. Взамен ГОСТ 7.1.—84, ГОСТ 7.16.—79 (нотные издания), ГОСТ 7.18.—81 (картографические издания), ГОСТ 7.34.—81 (изоиздания), ГОСТ 7.40.—82 (аудиовизуальные издания). — Введ. с 01.07.2004 г.
- 3) ГОСТ 7.9—95 (ИСО 21 4-76). Реферат и аннотация. Общие требования. — Взамен ГОСТ 7.9—77.— Введ. 01.07.97 г.
- 4) ГОСТ 7.66—92. (ИСО 5963—85). Индексирование документов. Общие требования к координатному индексированию. — Введ. 01.01.93 г.
- 5) ГОСТ 7.25—2001. Тезаурус информационно-поисковый одноязычный. Правила разработки, структура, состав, форма. Взамен ГОСТ 7.25—80.— Введ. 01.07.2002 г.
- 6) ГОСТ 7.82—2001. Межгосударственный стандарт. Библиографическая запись. Библиографическое описание электронных ресурсов. Общие требования и правила составления.— Введ. (впервые) 01.07.2002 г.

- 7) Воройский Ф.С., Острая С.А. Информационные ресурсы арбикона – пора подумать об их качестве // ”Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества”: Материалы конф. «КРЫМ-2004». М.:ГПНТБ России, 2004 г. <http://www.gpntb.ru/win/inter-events/crimea2004/disk/chapter2.html#section11/>.
- 8) Воройский Ф.С. Основные принципы обеспечения информационного поиска в корпоративных электронных каталогах // «Информационные технологии, компьютерные системы и издательская продукция для библиотек»: доклады и тезисы докладов МК «LIBCOM-2004».— М.: ГПНТБ России, 2004. — С. 56-59.