

СИСТЕМНЫЙ ПОДХОД К РАЗРАБОТКЕ ЛИНГВИСТИЧЕСКОГО ОБЕСПЕЧЕНИЯ АБИС

Соколовская Т.Б., ведущий библиотекарь

Харьковский национальный технический
университет сельского хозяйства им. П. Василенко

Автоматизация библиотечных процессов не только изменяет библиотечную технологию, но и создает основы для расширения информационных функций библиотек, повышения роли библиотек в формировании информационной среды научных исследований, библиотечно-информационного обеспечения организации учебного процесса, может создать большие возможности для изучения информационных потребностей пользователей.

Активное использование информационно-коммуникационных технологий приводит к трансформации всей библиотечно-библиографической деятельности, к необходимости профессиональной переориентации библиотечных работников [1].

Лингвистическое обеспечение автоматизированной библиотечно-информационной системы (АБИС) – комплекс мероприятий, направленных на разработку и эффективное использование языковых средств при сборе, обработке, распределении и выдаче информации пользователям, а также сами языковые средства.

Лингвистическое обеспечение АБИС включает в свой состав три направления деятельности библиотеки:

- 1) лингвистическое обеспечение подготовки баз данных;
- 2) лингвистическое обеспечение подготовки информационных изданий;

- 3) лингвистическое обеспечение процессов распределения потоков информации между пользователями и, в частности, процессов поиска.

Средства лингвистического обеспечения АБИС играют роль системообразующего фактора по отношению к другим обеспечивающим подсистемам АБИС. Эти средства образуют не только комплекс информационно-поисковых языков для индексирования документов и запросов, но и комплекс расширенных средств ввода информации в базу данных, средства анализа документальных информационных потоков.

Как пишет В. Г. Дригайло, АБИС в первую очередь должны удовлетворять требованиям системности и эффективности [2].

В основе разработки лингвистического обеспечения таких систем необходимо положить принципы функциональной полноты, простоты эксплуатации, достаточно высокие показатели полноты и точности информационного поиска.

В настоящее время корпоративное взаимодействие становится неотъемлемой составляющей инновационных процессов в библиотеках. Эти процессы порождают более жесткие требования к построению лингвистического обеспечения АБИС для поддержания лингвистической совместимости электронного каталога библиотеки с электронными каталогами других библиотек [3].

При проектировании лингвистического обеспечения АБИС приходится учитывать, что у этих систем три типа пользователей: сотрудники локальной библиотеки, библиотеки-партнеры по корпоративному проекту, читатели. У каждого типа пользователей свои требования к комфортности информационного поиска. Поэтому организация информационного поиска строится на основе трех типов моделей:

1. Модели внутрибиблиотечных процессов.

2. Модели корпоративного взаимодействия.

3. Модели взаимодействия библиотеки с читателями.

Эти три модели взаимосвязаны и взаимозависимы. Участие библиотеки в системе корпоративной каталогизации предполагает наличие подготовительного этапа, включающего:

- оптимизацию процессов ввода и редактирования информации;
- унификацию методики индексирования документов;
- реализацию возможности многократного и разнообразного использования однажды созданных средств лингвистического обеспечения.

Система ИРБИС интегрируется в современные корпоративные системы и технологии [5] не только благодаря совместимости с международными библиографическими форматами и поддержке Web-технологии и протокола Z39.50, но и благодаря наличию системы создания и ведения полного комплекса современных средств лингвистического обеспечения.

Система ИРБИС представляет собой открытую систему. Это означает возможности адаптации, модификации средств лингвистического обеспечения в зависимости от требований, соответствующих построенным функциональным моделям (1,2,3). Одно и то же средство может выполнять различные функции в зависимости от вида базы данных, от технологического процесса, который работает в определенной базе данных.

Система ИРБИС позволяет комбинированно применять разные средства лингвистического обеспечения для индексирования документов и запросов.

По технологии создания можно выделить такие типы лингвистических средств:

1. Создаются автоматически.
2. Создаются в автоматизированном режиме.
3. Создаются вручную.

В системе ИРБИС имеется возможность при необходимости изменять режим формирования тех или иных лингвистических средств.

Лингвистическое обеспечение процессов подготовки баз данных включает лингвистическое обеспечение процессов ввода и корректуры текстов. Эти средства можно разделить на две группы: 1) расширенные средства ввода данных, 2) средства корректуры данных. Расширенные средства ввода данных позволяют унифицировать форму вносимых данных, сэкономить трудозатраты на ввод данных. К таким средствам относятся формируемые вручную справочники: «Тип документа», «Вид документа», «Характер документа», «Место хранения документа».

Особую группу расширенных средств ввода данных представляют формируемые в автоматизированном режиме авторитетные файлы. Система ИРБИС позволяет средствами глобальной пакетной обработки файла электронного каталога создавать и пополнять базы данных авторитетных файлов (в нашем случае - авторов, АПУ к УДК, АПУ к ББК) с последующей их корректировкой вручную.

В системе ИРБИС имеются возможности создания иерархических рубрикаторов, которые можно использовать и как информационно-поисковый язык (ИПЯ), и как расширенное средство ввода данных, например, рубрикатор подразделений ХНТУСХ, который функционирует в виде иерархического дерева.

Интерес представляют разработанные библиотекой ХНТУСХ средства автоматизированного группового ввода данных, позволяющие при вводе данных в одно поле/подполе автоматически производить ввод в связанные семантически с этим полем/подполем поля/подполя записи библиографического описания документа. Например, при вводе из авторитетного файла авторов-сотрудников ХНТУСХ ФИО сотрудников в базу

данных «Труды сотрудников ХНТУСХ» автоматически вносятся коды подразделений и признак работы автора в ХНТУСХ; при вводе данных об авторах из автоматически создаваемого словаря в базе данных «Электронный каталог статей» одновременно заполняются подполя: ФИО автора, ученая степень, должность.

Эффективным средством групповой корректуры данных является корректировка непосредственно по автоматически создаваемому словарю без обращения к отдельным записям базы данных. По словарю легко выявляются ошибки ввода или индексирования данных, однотипные ошибки устраняются непосредственно при работе в режиме «Словарь» одной командой пакетной корректировки сразу всех записей базы данных, содержащих одинаковые ошибки. Это особенно важно при корректуре конвертированных баз данных.

Система ИРБИС содержит средства анализа базы данных, благодаря которым можно выявить наличие или отсутствие определенного элемента данных в записях, произвести поиск записей с отсутствующими или ошибочными элементами и затем произвести глобальную корректировку записей в пакетном режиме.

Системность средств лингвистического обеспечения приводит к взаимозависимости процессов ввода, поиска и вывода информации. Понимание этой взаимозависимости необходимо для эффективной работы системы во всех режимах. Поэтому в библиотеке Харьковского национального технического университета сельского хозяйства им. П. Василенко (ХНТУСХ) было проведено специальное занятие с сотрудниками отдела комплектования и научной обработки литературы по принципам информационного поиска в системе ИРБИС. На занятии рассматривались вопросы:

- 1) Что влияет на качество информационного поиска.
- 2) Поиск в ИРБИС на полноту и точность.
- 3) Виды информационного поиска в системе ИРБИС.

Во время занятия было наглядно показано, как ошибки ввода непосредственно влияют на формирование автоматически создаваемого словаря по различным полям. Уделено было специальное внимание тем ошибкам, которые трудно выявить при непосредственном редактировании отдельной записи. Такие ошибки легко определить при анализе словаря. Например, наличие пробелов в индексах УДК, ББК, первые пробелы, знаки препинания перед ключевыми словами, инвентарными номерами и т.п. Ключи с подобными ошибками система помещает вне основного ряда сортировки и при формировании поискового запроса они в запрос не попадают.

На занятии были рассмотрены методы быстрого поиска записей, нуждающихся в редактировании. Кроме того, было показано, какими методами администратор базы данных может производить быстрые корректировки в больших массивах данных (глобальные корректировки). Администратор баз данных рассказал о том, что интегральность системы ИРБИС означает возможность при однократном правильном вводе данных выводить их разными способами в разных комбинациях, приписывая данным из определенного поля/подполя какой-то текст (если это необходимо).

Каждый тип пользователей нуждается в общении с системой на привычном ему языке. Для библиотечных работников привычным является язык классификационных индексов (УДК, ББК) и язык библиографических данных (например: авторы, наименования издательств, ключевые слова заглавия). Ученым и преподавателям гораздо удобнее общаться с системой на языке тематических и предметных рубрик.

Совокупность средств лингвистического обеспечения по мере формирования баз данных позволяет построить модель политематической предметной области, отражающей структуру информационных потребностей

разных категорий пользователей, а также структуру деятельности университета как учебно-научной организации.

Иерархический рубрикатор подразделений ХНТУСХ отражает структуру деятельности университета. Он используется в базе данных «Труды сотрудников ХНТУСХ» для обеспечения выборок по подразделениям, кафедрам университета, а в базе данных «Электронный каталог книг» этот рубрикатор позволяет выбирать информацию для информирования кафедр о новых поступлениях в библиотеку и готовить информационные листы в соответствии с запросами каждой кафедры. Рубрикатор представлен в виде дерева. Такая структура позволяет быстро найти информацию по нужному подразделению или кафедре.

Лингвистическое обеспечение в полноценной АБИС должно реализовывать тематический поиск [3]. Такой поиск поддерживается различными классификационными системами. При этом гибкость классификационной системы может иметь не только положительную, но и отрицательную сторону. Поэтому возникает необходимость регламентирования процессов индексирования документов средствами автоматизации. Это особенно важно при разработке лингвистического обеспечения корпоративных систем, так как позволяет обеспечить унифицированную методику индексирования для библиотек-партнеров.

Интерес представляют разработанные библиотекой ХНТУСХ авторитетные файлы «АПУ УДК» и «АПУ ББК», которые обеспечивают возможность составления поискового запроса для осуществления тематического поиска в удобной для пользователей форме. Использование только свободных ключевых слов без учета смысловых связей между ними обуславливает большие потери информации, невидимые пользователю [6]. Авторитетные файлы «АПУ УДК» и «АПУ ББК»

позволяют формировать поисковый запрос по словесным формулировкам индексов – поиск по отдельным словам и их частям, имеющимся в словесных формулировках индексов. В АРМ «Читатель» имеется специальный режим, который позволяет осуществлять поиск по авторитетным файлам. В АРМ «Каталогизатор» есть возможность использовать авторитетные файлы непосредственно при индексировании документов и составлении библиографической записи.

В режиме «Словарь» производится поиск нужной формулировки по ключу, а затем можно просмотреть полностью нормативную запись авторитетного файла и выбрать для поиска связанные рубрики.

Создание авторитетного файла «АПУ УДК» - начальный этап для построения иерархической базы данных УДК, которая дает возможность составлять запросы по иерархическому дереву и также может быть подключена непосредственно к режиму ввода информации в АРМ «Каталогизатор». Автоматизированное построение авторитетного файла «АПУ УДК» позволяет выбрать из электронного каталога рубрики, сформулированные каталогизаторами библиотеки. Затем эти рубрики переносятся в иерархическую базу данных «УДК», где можно поместить комментарий по методике отнесения документов к данной рубрике. В иерархической базе данных УДК реализуются функции АПУ: поиск нужной рубрики по ключевым словам, а затем возможен поиск по найденной рубрике в электронном каталоге.

Все перечисленные средства могут применяться в комплексе с автоматически создаваемым в ИРБИС словарем.

Таким образом, средства лингвистического обеспечения позволяют использовать АБИС как инструмент по совершенствованию и повышению эффективности библиотечных технологий при работе с

базами данных в среде ИРБИС и как основу для корпоративного взаимодействия библиотек.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Чучукалова, Р.А. Информационная компетентность библиотекаря в системе информационного пространства Сибирского федерального университета [Текст] / Р.А. Чучукалова // Науч. и техн. б-ки. – 2010. - № 4. – С.58-62.
2. Дригайло, В.Г. Технология работы библиотеки [Текст] /В.Г. Дригайло. – М.: Либерей-Бибинформ, 2009. – 544 с.
3. Зайцева, Е.М. Лингвистическое обеспечение автоматизированных библиотечно-информационных систем: современные требования и направления развития [Текст] / Е.М. Зайцева // Науч. и техн. б-ки. – 2000. - № 3. – С.54-57.
4. Маторин, С.И. Системология и объектно-ориентированный подход [Текст] / С.И. Маторин // НТИ. Сер. 2. – 2001. - № 8. – С. 1-8.
5. Бродовский, А. Новая идеология ИРБИС [Электронный ресурс] // Первое сентября. – 2009. - № 13. – Режим доступа: <http://www.1september.ru>
6. Лавренова, О.А. Есть такая запись в электронном каталоге! [Текст] / О.А. Лавренова // Библиотека. – 2000. - № 2. – С. 30 – 33.