

РАЗРАБОТКА НАЦИОНАЛЬНОЙ БАЗЫ АВТОРИТЕТНЫХ ДАННЫХ В БЕЛАРУСИ

Необходимым условием внедрения автоматизированных библиотечных технологий сегодня является следование международным стандартам и рекомендациям. Это обеспечивает качественный уровень технологии и способствует интеграции библиотек в мировое информационное пространство. В области машиночитаемой каталогизации и создания библиографических ресурсов таким международным стандартом являются MARC-форматы библиографических записей. MARC-форматы имеют целый ряд особенностей, которые не свойственны локальным форматам автоматизированных библиотечно-информационных систем предыдущих поколений, получивших в наших библиотеках в прежние годы широкое распространение и используемых до сих пор. К числу отличительных особенностей MARC-форматов относятся, например, достаточно большой репертуар полей кодируемых данных, а также возможность выделения неcodируемых поисковых элементов библиографических записей в отдельную базу данных — базу авторитетных данных. Авторитетные данные включают поисковые элементы, форму наименования которых необходимо нормировать и контролировать для обеспечения высокой полноты и точности поиска в библиографических базах данных, а также для более качественного учета документопотока. В силу своей многочисленности и сложности соотношения с именуемыми объектами (или энтитетами) эти нормируемые наименования не могут и не должны заменяться кодами. К именам такого типа относятся имена лиц, коллективов, родовых кланов (династий), торговых марок, географические и административно-территориальные наименования, унифицированные заглавия анонимных классических произведений, лексические единицы вербальных информационно-поисковых языков, в том числе наименования видов, жанров, физических характеристик документов. База авторитетных данных позволяет учесть сложные соотношения между объектом реального мира и его одним или несколькими именами, а также между именами разных, но связанных в реальности объектов. Один и тот же объект может иметь несколько форм своего имени и даже несколько разных имен. Например, нередко лицо на протяжении своей жизни меняет свое имя по разным причинам (при выходе замуж, принимая духовное имя и т.д.) или пользуется псевдонимом в одной сфере деятельности, не отказываясь от своего подлинного имени -- в другой. Кроме того, имя одного и того же лица по недосмотру издателей может быть оформлено с искажениями при опубликовании разных его работ в разное время. Многие организации, территориально-административные образования меняют по несколько раз свои наименования на протяжении истории своего существования. Понятно, что наличие нескольких имен для одного объекта создает немало проблем в поиске информации, так как пользователь должен знать все возможные имена искомого объекта, чтобы получить требуемую полноту ответа на свой запрос. Это, как правило, нереально для большинства пользователей. С другой стороны, разные объекты могут иметь одинаковые имена. Например, нередко случаи совпадения фамилий, имен и отчеств у разных лиц, существуют одинаковые наименования населенных пунктов, слова-омонимы в составе вербальных ИПЯ. Неразграничение объектов с совпадающими именами создает большой поисковый “шум”, т.е. система выдает на запрос пользователя много лишних документов. Каталогизаторы, работавшие в традиционном режиме, давно пытались хотя бы в какой-то мере снимать эти негативные явления при формировании заголовков библиографических записей и используя ссылочно-справочный аппарат каталогов. Однако карточная форма каталогов позволяла делать это в ограниченной мере.

Базы авторитетных данных (или авторитетные/нормативные файлы в другой терминологии) решают непростую проблему соотношения имен, с одной стороны, путем однозначной идентификации самого объекта именования и установления для него принятой, “авторитетной” формы имени, а с другой — путем связывания в базе данных логическими связями разных форм имени объекта (как вариантов одного имени объекта, так и разных его имен). База авторитетных данных позволяет дифференцированно подходить к использованию разных форм имени объекта в информационном запросе. Некоторые формы имени, указанные пользователем в запросе, автоматически заменяются на принятую, установленную форму, используемую каталогизаторами; другие формы могут автоматически дополнять запрос пользователя (вышестоящие термины по отношению к термину, указанному пользователем, или параллельные формы имени на разных

языках); третьи формы могут быть включены в запрос только самим пользователем (например, исторически сменявшие друг друга формы наименования одной организации).

Установленные, принятые формы наименований объектов выносятся в заголовок основной и добавочных библиографических записей при формировании библиографических указателей, списков и других выходных форм для записей библиографических баз данных.

Базы авторитетных данных не являются строго необходимыми при использовании MARC-форматов, но из всего вышеизложенного понятно, что они очень полезны для процессов каталогизации, поиска, учета документопотока, потому что помогают более качественно выполнять эти технологические процессы.

Дифференциация имен и идентификация объектов (энтитетов), которым ставится в соответствие установленная, принятая, “авторитетная” форма имени, получающая в базе данных уникальный идентификатор, делают необходимым наличие определенного объема справочных сведений в каждой записи, сформированной для установленной формы имени. Это могут быть сведения как о самом объекте, так и об его имени. Наличие справочных сведений в авторитетных записях делает базу авторитетных данных самоценным информационным ресурсом, пригодным и для информационной, справочной работы библиотек, архивов, музеев, а также позволяет считать такую БД одной из форм сохранения культурного наследия народа.

ИФЛА придает большое значение созданию национальных баз авторитетных данных библиографирующими центрами стран мира. Национальные авторитетные файлы должны регистрироваться ИФЛА и выставляться в Internet для всеобщего ознакомления. Под эгидой ИФЛА и на основе UNIMARC был разработан формат для авторитетных записей UNIMARC/AUTHORITIES, опубликованный в 1991 г. Вторая версия этого формата, существенно доработанная, появилась в 2001 г.

В Беларуси освоение технологии авторитетного контроля и создания авторитетных файлов началось в 1998 г., параллельно с разработкой формата BELMARC. В 1999—2000 гг. была разработана первая версия формата для авторитетных записей BELMARC/AUTHORITIES на основе версии UNIMARC/AUTHORITIES 1991 г. С 2001 г. в республике в рамках государственной Программы информатизации на 2001—2005 гг. проводятся работы по проектированию автоматизированной системы создания и ведения национальной БД авторитетных записей. Тема выполняется Национальной библиотекой Беларуси, соисполнителями выступают Национальный центр информационных ресурсов и технологий РБ (в системе НАН Беларуси) и НПФ “ИНЕАК”. Руководителем темы является зам. директора НББ Т.В.Аксенова. При Национальной библиотеке Беларуси создан творческий коллектив (руководитель — доцент, автор данной статьи С.В.Ильина), в состав которого входят ведущие специалисты НББ, НКП Беларуси, некоторых крупнейших библиотек г. Минска. В рамках проекта дорабатывается формат авторитетных записей с учетом изменений UNIMARC/AUTHORITIES в версии 2001 г., конвертируются машинные словари НББ в структуру авторитетных файлов, разрабатывается технология формирования авторитетных записей в процессе каталогизации литературы и технология ведения национальной базы авторитетных данных, создается программное обеспечение подготовки и использования авторитетных записей.

Национальная база данных авторитетных записей должна включать наименования национальных объектов (энтитетов): лиц, коллективов, географических объектов, книжных серий, торговых марок, а также лексические единицы вербальных информационно-поисковых языков, разработанных в Беларуси. Она будет насчитывать несколько десятков тысяч, а в перспективе — миллионы записей. Это действительно корпоративный ресурс, который может быть создан только совместными усилиями многих библиотек Беларуси. Видимо, в будущем к этой работе подключатся музеи и архивы.

В процессах каталогизации документов национальная БД авторитетных записей будет использоваться не изолированно, а в составе более общего комплекса баз авторитетных данных, поддерживающих библиографические ресурсы, генерируемые библиотеками Беларуси. Записи национальной БД в этом комплексе будут отмечены принадлежностью Беларуси отражаемых в них энтитетов, а в случае вербальных ИПЯ в записях на каждую лексическую единицу будет внесено условное обозначение соответствующего словаря, разработанного в Беларуси.

Технология создания и использования авторитетных данных предполагает программный перенос по команде каталогизатора установленной формы имени энтитета и ее уникального идентификатора из базы авторитетных данных в соответствующее поле точек доступа (поисковое

поле) библиографической записи. В случае отсутствия требуемого имени в БД авторитетных записей сначала должна быть создана авторитетная запись, а затем путем программного переноса сформирована точка доступа библиографической записи.

В процессе проектирования белорусским разработчикам приходится решать ряд достаточно сложных вопросов, обусловленных объективными причинами.

Во-первых, международный стандарт формата для авторитетных записей имеет сравнительно недолгую историю, поэтому он постоянно совершенствуется, но изменения в него вносятся редко из-за довольно сложной процедуры их принятия Постоянным комитетом по UNIMARC при ИФЛА. Поэтому до сих пор, даже в значительно улучшенной версии 2001 г., в тексте этого документа имеются не очень четко сформулированные рекомендации, сопровождаемые недостаточно полными примерами. Это создает неоднозначность толкования некоторых существенных положений формата. Во-вторых, форматы UNIMARC — это обменные форматы, в задачу которых входит стыковка различных локальных форматов с целью обмена данными. Форматы BELMARC проектируются путем адаптации UNIMARC, в первую очередь, с целью обеспечения каталогизации и создания логических схем интегрированных библиографических баз данных, генерируемых в Беларуси. Однако наши правила каталогизации по ГОСТ 7.1-84 далеки от англо-американских правил AACR2 и далеко не полностью гармонизированы с международными рекомендациями ISBD, а, как известно, именно на ISBD и опираются форматы UNIMARC. И хотя как раз сейчас наши правила каталогизации пересматриваются с целью их максимального сближения с ISBD, оформленных решений пока нет. В-третьих, традиции формирования заголовка библиографической записи у нас существенно иные, чем во многих зарубежных странах. Для нас неприемлемы их подходы к оформлению установленной, принятой формы заголовка в авторитетных записях. В-четвертых, белорусские библиотеки и НКПБ пользуются не одним языком каталогизации, как большинство зарубежных библиографирующих центров, а несколькими языками. А это, в свою очередь, напрямую влияет на выбор языка представления установленной формы имени объекта, которая выносится в заголовок библиографической записи.

На международном уровне пока еще только обсуждается проблема создания многоязычных авторитетных баз данных, в которых на равных началах использовались бы установленные формы имен на разных языках в составе одной базы данных. Для Беларуси это чисто практический вопрос по причине двуязычия и укоренившейся традиции формировать заголовок библиографической записи на языке библиографического описания. Если мы считаем, что белорусскоязычная или русскоязычная форма имени одного и того же энтитета для наших пользователей имеет равное значение и, значит, любая из них может быть вынесена в заголовок библиографической записи в зависимости от языка текста каталогизируемого документа и языка библиографического описания, то обе эти формы должны считаться нормативными, или установленными, принятыми. Но в поиске обе эти формы должны учитываться совместно, что обычно в других странах применяется для вариантных, а не связанных установленных форм имени.

Приведенный далеко не полный перечень спорных вопросов показывает, насколько непроста адаптация международных рекомендаций по форматам к национальным условиям их использования. Поэтому некоторые из проектных решений, принимаемых сегодня, в будущем придется пересмотреть. Однако спорные, трудно решаемые проблемы не должны останавливать проектные и практические работы по созданию баз данных авторитетных записей.

Белорусские библиотеки давно уже осознали необходимость авторитетного контроля и создания баз авторитетных данных. Ими накоплен определенный материал для этой работы, что показало исследование, проведенное рабочей группой в 1998 г. Есть и реальные эксперименты по созданию прототипов машиночитаемых авторитетных данных, о чем свидетельствует опыт Президентской библиотеки. Однако следует заметить, что большинство пакетов прикладных программ для АБИС, используемых библиотеками Беларуси сегодня, не позволяют в полной мере применять технологию авторитетного контроля, так как при разработке этих программных комплексов такая технология не закладывалась. Однако независимо от возможностей своих локальных систем многие библиотеки могли бы принять участие в формировании авторитетных записей, если будет существовать принятый этими библиотеками порядок по данному направлению и создана единая служба поддержки авторитетных баз данных в республике.