

Библиотечные стандарты и стандартизация в области информационно-поисковых систем: состояние дел и направления дальнейших исследований

Б.Р. Логинов

ООО "ДИТ-М" (Документальные информационные технологии)

Адрес: 109240, Москва, Москворецкая наб., 2а 1 тел. 298 59 29

факс 925 96 50

E-mail: loginov@ditm.ru

<http://www.ditm.ru>

И.Б. Цветкова

Общепризнанно, что стандартизация в области информационно-поисковых систем (ИПС) и особенно в области языков индексирования является важнейшим условием эффективной организации процессов обработки документов и их поиска, обеспечения качества библиографической информации и возможности обмена ею.

Осознание необходимости и преимуществ стандартизации в библиотечной среде пришло с началом использования машиночитаемых библиографических записей (БЗ) и обмена ими. С этим периодом связано и начало разработки отечественной системы стандартов по информации, библиотечному и издательскому делу – системы СИБИД. Своего рода “золотой век” библиотечной стандартизации пришелся на конец 70-х – начало 80-х гг. В это время появляется порядка 30 ГОСТов, непосредственно относящихся к работе библиотечно-библиографических служб. Но, к сожалению, затем темпы работ постепенно снизились, и в последние 15 лет появление новых библиотечных стандартов, скорее, исключение, чем правило.

Большинство так называемых ГОСТов серии 7, относящиеся к исследуемой нами предметно-тематической области, были разработаны как терминологические словари, отражающие уровень тогдашних научных представлений о способах организации, ведения и редактирования ИПС, а также о самих информационно-поисковых системах и языках индексирования, используемых для их организации. Исключение составляют ГОСТы на библиографическое описание (БО), устанавливающие общие принципы и правила его составления для различных видов документов: набор элементов БО, последовательность их расположения, наполнение и способ представления каждого элемента, применение условных разделительных знаков.

Стандарты на БО появились как результат пересмотра национальных правил каталогизации в связи с созданием в 1974 году Международного стандартного

библиографического описания (ISBD). Как известно, ISBD превратилось в международный стандарт ISO, который требовал пересмотра национальных правил в странах-членах ISO и создания соответствующих национальных стандартов.

Создание ГОСТов на БО, безусловно, - наше главное достижение в области стандартизации. Это стало возможным, прежде всего благодаря успехам описательной каталогизации, которая всегда была в нашей стране одной из наиболее развитых областей библиотечно-библиографической теории и практики, важнейшей тенденцией развития которой стали унификация и стандартизация, как на национальном, так и на международном уровне.

Тем не менее, давая общую оценку состояния отечественной стандартизации в области ИПС, следует отметить, что терминология большинства стандартов устарела, их система фрагментарна и не охватывает всего комплекса объектов и процессов, требующих стандартизации.

Так, любой, кто сегодня приступает, например, к созданию электронного каталога, не сможет найти в нашей системе нормативных документов необходимого набора стандартов, руководств и правил, объясняющих и регламентирующий этот процесс (за исключением ГОСТов на БО и национальных правил каталогизации, которые тоже подлежат пересмотру под мощным воздействием новой электронной среды).

Это можно объяснить, прежде всего, тем, что отечественная теория библиотечно-библиографических ИПС, являющаяся базой стандартизации, развивается очень медленно и не отражает происходящих на практике качественных изменений и, прежде всего, в способах создания и распространения информации, связанных с переходом на новые информационные технологии.

Электронная форма каталога и возможности использования телекоммуникаций для создания и обмена БЗ привели к переменам, заставляющим не просто пересмотреть основополагающие принципы теории и практики каталогизации или организации и ведения библиотечно-библиографических ИПС, а в большинстве случаев практически заново их создать.

Определим главные “лакуны” в теории и методике библиотечно-библиографических ИПС. Два явления, два термина пошатнули их устои – “электронный каталог” и “машиночитаемая каталогизация”.

Отечественная теория и практика библиотечно-библиографических ИПС до последнего времени разрабатывалась и развивалась, прежде всего, применительно к карточным каталогам. Традиционная типология делила их по видам в зависимости от принципов группировки БЗ и типов запросов. Электронные каталоги выделялись как форма каталога (наряду, например, с печатной или на микрофишах).

Но вот вместо множества (системы каталогов) мы сегодня получили один каталог - электронный. ЭК предлагает все варианты поиска вместе, отвечает на все типы

запросов, хотя и не группирует записи внутри себя по какому-либо признаку. К какой категории его отнести – формы или вида? В отечественной теории и практике до сих пор не существует сколько-нибудь научного определения ЭК. Достаточно привести определение, данное ему в последнем по времени издания (1997) терминологическом словаре по библиотечному делу: “библиотечный каталог в машиночитаемой форме, работающий в режиме реального времени, предоставленный в распоряжение читателей библиотеки. Работа с ЭК производится с помощью терминала”.

В зарубежной теории принято выделять в структуре ЭК три основных части: библиографический файл или файл БО, авторитетные файлы (индивидуальных имен и наименований коллективов, унифицированных заглавий, предметных рубрик и классификационных индексов), а также файл сведений о физическом экземпляре.

Создание двух взаимосвязанных файлов в ЭК: библиографического и авторитетного, стало возможным благодаря концепции разграничения БО и заголовка (поисковых признаков, точек доступа), получившей развитие в международной теории и практике каталогизации.

Характерно, что ISBD, на котором базируются наши стандарты БО, уже в 70-е годы предусматривало определенную “машиночитаемость”, так как создавалось, когда уже существовал и успешно использовался формат MARC. Структура ISBD была изначально ориентирована на электронную среду: библиографическая информация была дискретной, с разметкой для дальнейшего использования в машиночитаемом формате. Оно уже тогда не включало поисковые признаки: заголовки основной и добавочной записи, что создавало условия для организации новой структуры каталога с организацией библиографического и авторитетного файлов. К сожалению, идея разделения заголовка и БО долгое время не признавалась в отечественной теории каталогизации, и не была учтена при создании наших ГОСТов. Кроме этого в большинстве библиографирующих учреждений отсутствовала практика ведения авторитетного контроля и ведения специальных авторитетных файлов (АФ).

За рубежом ведение авторитетного контроля и авторитетных справочников и картотек имеет долгую и богатую историю. Уже в 19 веке основные принципы авторитетного контроля были определены Ч.Кеттером. Публикации 20 века отражают весь диапазон исследований и взглядов на авторитетный контроль и АФ, а с начала 70-х годов в связи с внедрением новых технологий он неизменно выделяется как самостоятельное понятие в теории и самостоятельный технологический процесс в практике ведения каталогов.

В наши дни действуют многочисленные международные проекты в области обмена АФ, а также в области создания международных АФ.

Тем не менее, у нас эта проблема до сих пор остается мало изученной. Так, в новом ГОСТе системы СИБИБД, введенном в 1998 году, 7.76-96, в разделе “Заголовки библиографической записи” вместо термина авторитетный файл введен термин “официальный справочный перечень”. Он определяется как “

перечень авторитетных (специально установленных и признанных нормативными) заголовков, используемых в каталоге, например, имен лиц, наименований организаций, заглавий анонимных классических произведений, предметных рубрик, дескрипторов”.

В то же время, нельзя не отметить и определенные успехи: в рамках программы ЛИБНЕТ создан формат для обмена авторитетными данными, и активно разворачиваются работы по созданию системы национальных авторитетных файлов.

А вот как определяет уже цитируемый нами терминологический словарь по библиотечному делу машиночитаемую каталогизацию: “ Система каталогизации, используемая для организации и распространения в машиночитаемой форме структурированных в соответствии с международными правилами каталогизации библиографических данных, объединенных в национальные и локальные базы данных”. При этом просто “каталогизация” определяется как “совокупность процессов, обеспечивающих создание и функционирование библиотечных каталогов. К ним относятся формирование БЗ (составление БО, систематизация, предметизация), ввод данных в автоматизированную информационную систему или тиражирование библиотечных карточек, работу с каталогами (организация, ведение и редактирование), управление системой каталогов и технологическими процессами каталогизации”. Таким образом, традиционная каталогизация – это процесс, а машиночитаемая – специальная система для организации библиографических данных.

Попробуем все же, как и в случае с ЭК, рассмотреть “машиночитаемую каталогизацию” с позиции вида. Это новый вид каталогизации? Чем тогда он отличается от традиционной, предназначенной для карточных каталогов? Тогда соответственно должны существовать машиночитаемая предметизация и машиночитаемая систематизация?

Нам кажется более правильным все же определять каталогизацию как процесс составления каталога. Результатом этого процесса может стать карточный, печатный, электронный и любой другой каталог.

Рискнем также предположить, что термин машиночитаемая каталогизация (от английского MARC –Machine Readable Cataloguing) принадлежит к разряду ложно ориентирующих и на самом деле отражает не новый вид каталогизации, а новую технологию создания машиночитаемой БЗ с использованием специального формата представления ее элементов. Соответственно, MARC-формат – это инструмент, логическая структура для внутреннего представления элементов БЗ в базе данных и правила их создания. Принято различать формат каталогизации и коммуникативный. Вполне вероятно, что в будущем элементы машиночитаемых БЗ смогут быть представлены с использованием других инструментов, а MARC-форматы, как и каталожные карточки, уйдут в прошлое. Но, надеемся, что это – далекое будущее. А сегодня в России завершается создание системы форматов представления БЗ в машиночитаемой форме.

Вынуждены отметить, что в работе над созданием системы форматов RUSMARC,

тоже не удалось избежать ошибок в применении термина “машиночитаемая каталогизация”. Один из форматов назван форматом машиночитаемой каталогизации, тогда как следовало назвать его форматом представления машиночитаемых библиографических записей.

Отдельно следует остановиться на стандартах содержательного анализа документов. В отличие от описательной каталогизации, систематизация и предметизация (за исключением терминологии) не были до сих пор представлены в системе отечественных стандартов.

Процесс содержательного анализа документов включает два четко выделяемых этапа - интеллектуальный анализ документа и приписывание индексов, предметных рубрик или дескрипторов. В литературе почти не рассматривается первый - интеллектуальный этап. Большинство работ о нем носят теоретический или экспериментальный характер, практических "стандартов" или "руководств" для этого аспекта процесса индексирования не создано. Действительно, достичь стандартизации процесса содержательного анализа было бы трудно и вряд ли возможно вообще. Однако жизненно важное значение имеют "стандарты" и "руководства" по разработке самих языков индексирования и источники таких стандартов.

Системы классификации как стандарты.

В ответ на запрос "классификация" и "стандарты" компьютером было найдено 165 записей. Большая часть этих записей касается не стандартов классификации как таковых, а официального или неофициального принятия одной или нескольких систем классификации в качестве стандартов. То есть системы классификации сами по себе являются принятыми стандартами; указания на существование "стандартных правил" разработки классификации найти трудно, хотя существуют основные принципы классификации, которые должны использоваться при разработке подобных систем.

Наиболее известные современные системы классификаций были разработаны либо крупными библиотеками, либо специальными комитетами, и носят международный характер. Это верно и для Десятичной Классификации Дьюи, Универсальной Десятичной классификации (УДК), и для ББК в России. Так как в основе любой системы классификации лежит особая, специфическая философия и любая система классификации имеет ряд уникальных свойств, неудивительно, что не разработано общих стандартов или руководств по созданию таких систем. Тем не менее, принять стандарты в этой области необходимо.

Вербальные языки и стандарты.

Среди систем вербальных языков обычно выделяют два типа - системы рядов терминов, часто называемых предметными рубриками и разработанными для предкоординированных систем, и системы понятий, или дескрипторов, называемых тезаурусами и разработанными для систем с посткоординацией.

Стандартизация контролируемых словарей осуществляется двумя способами -

через совместное использование списков стандартных заголовков предметных рубрик на национальной или международной основе и через применение комплекса специальных руководств. Примером стандартизации в этой области являются такие средства, как Список предметных рубрик Библиотеки Конгресса и подобные ему списки предметных рубрик в Германии, Франции, Финляндии и Польше. Многие такие списки по праву стали "стандартами" благодаря совместному использованию и, кроме того, применению в машиночитаемых массивах, которые широко распространяются среди информационных агентств.

Возвращаясь к характеристике системы библиотечных стандартов в области ИПС, отметим, что она не должна существовать только как набор терминологических ГОСТов. Целесообразно расширить ее за счет включения важнейших достижений международной стандартизации: форматов MARC, национальных авторитетных файлов и их источников.

Как сделать их стандартами? В первую очередь, попытаемся разобраться в том, что представляют собой стандарты в более широком, чем терминологические ГОСТы, плане.

В наиболее формальном смысле стандарт - это набор правил, установленных неким полномочным органом и предписанных для выполнения организациями, агентствами, институтами, на национальной или международной основе. В этом смысле стандарты имеют обязательный характер, и их выполнение может поддерживаться правительством или административными органами, включая различные профессиональные ассоциации. Например, правительство может законодательно закрепить стандарты здоровья для населения страны, в то время как ассоциации врачей могут устанавливать стандарты этики в качестве основы дисциплины членов этих ассоциаций, выполнения ими профессиональных обязанностей на достаточно высоком уровне и соблюдения определенных норм поведения. Этот тип стандарта необходим в тех случаях, когда недостаток стандарта может привести к угрозе жизни или к ситуации хаоса. Под эту категорию подпадают очень немногие (если вообще подпадают) из библиотечных стандартов. С достаточной долей условности к этой группе стандартов можно отнести наши ГОСТы. (Но, тем не менее, трудно себе представить, что авторитетный файл имен индивидуального автора станет ГОСТом).

Также формальными, но несколько менее регулируемыми являются стандарты, которые описываются в зарубежной литературе как "руководства" или "правила". В общем смысле такие правила вырабатываются группами авторитетных экспертов и принимаются в качестве "желательной" основы действий и их оценок. Слово "руководство" предполагает, что правила являются не обязательными, а скорее рекомендательными, факультативными. В связи с этим в некоторых случаях они будут применены, в других - нет. В библиотечном мире, как нам представляется, существует тенденция к разработке не строгих обязательных стандартов, а именно правил такого типа, позволяющим каждой организации и институту самим решать, применять ли такое руководство, и если применять - то полностью или только частично.

К этой группе стандартов можно отнести национальные правила каталогизации,

общие и частные методики систематизации и предметизации документов, а также правила создания и ведения авторитетных файлов (которые предстоит еще создать).

Третий, наименее формальный подход к стандартизации - это подход через использование всеми одних и тех же средств и инструментов создания и распространения БЗ. Т.е. средство, например список предметных рубрик, принимается в качестве нормы и "по умолчанию" становится стандартом. Стандарты этого типа вырабатываются не административным органом или группой экспертов, а являются результатом согласованности действий.

Большинство библиотечных стандартов берут начало в простых, чисто практических начинаниях. Часто они начинаются с политики отдельных библиотек, в которых устанавливаются некоторые правила и инструкции для того, чтобы способствовать эффективной деятельности в конкретных, специфических условиях. Если одни и те же правила и инструкции принимаются не одной библиотекой, можно сказать, что библиотеки, использующие одни и те же правила, выработали "стандарты" или "инструкции" для данной группы библиотек. По мере того, как эти правила и процедуры принимаются другими институтами, влияние "стандартов" возрастает. В конечном итоге они могут стать национальными стандартами и со временем оказать влияние на развитие международных стандартов. Именно такой тип развития можно наблюдать в истории правил каталогизации: от Правил Британского Музея до Англо-американских Правил каталогизации. Использование предметных заголовков Библиотеки Конгресса на национальном и международном уровне также сделало этот словарь стандартом.

К этой группе могли бы быть отнесены форматы, авторитетные файлы, национальные системы классификации, списки предметных рубрик и т.п.

Все вместе эти три группы и должны составить систему библиотечных стандартов в области ИПС. Ее, как таблицу Менделеева, предстоит заполнить конкретными документами и начать работу. Целесообразно для этого образовать специальную комиссию или программу.

Конечно, мы отдаем себе отчет, что простое стремление к качеству или развитие теоретических основ не приводит к стандарту. Для того чтобы появились стандарты, должна существовать осознанная и настоятельная необходимость в них. Как правило, эта необходимость применительно к нашей области исследования определяется практическими вопросами.

Логинов Борис Родионович - родился в 1946 г. Закончил МИФИ в 1970 г. по специальности прикладная математика. В 1975 г. защитил кандидатскую диссертацию в аспирантуре МИФИ по созданию информационной технологии автоматизированного проектирования в АСУ. Работал преподавателем МИФИ до 1978 г., руководителем отдела математического моделирования ВНИИ СГ и ОЗ им. Н. А. Семашко до 1986 г., зам. ген. директора НПО "Союзмединформ". Директор

Государственной центральной научной медицинской библиотеки с 1986 г. по настоящее время. С 1994 г. Президент научно-внедренческой компании “ДИТ-М”, занимающейся разработкой автоматизированных информационных технологий для библиотек. Доцент Московского государственного университета культуры на библиотечном факультете.

Общественная работа: Вице-президент Российской библиотечной ассоциации, научный руководитель федеральной программы “Создание российской компьютерной библиотечной сети ЛИБНЕТ” Министерства культуры РФ, председатель секции “Информационные технологии в медицине” Министерства здравоохранения РФ.

Автор более 100 научных работ. Один из создателей автоматизированных библиотечных систем ДИТ-ИБИС, OPAC-R, OPAC-Global.

Цветкова Ирина Борисовна - канд. пед. наук, руководитель программы по созданию российских форматов представления БЗ в машиночитаемой форме (RUSMARC)

Компания ДИТ-М является одной из ведущих компаний на российском рынке по производству программных продуктов для автоматизации библиотек и библиотечных сетей. В 1989 году компания впервые вышла на рынок с системой ДИТ-ИБИС, которая, развиваясь в течение шести лет на основе ППП CDS ISIS для ОС DOS, к 1994 году превратилась в полномасштабную сетевую систему, включающую автоматизацию всех основных библиотечных процессов: комплектование с подпиской, межбиблиотечный книгообмен, каталогизацию, индексирование с помощью встроенного тезауруса, поиск с использованием тезауруса, поддержка выходных форм в виде гостированных библиографических карточек (ГОСТ 7.1-84), заказ и книговыдачу, ведение файлов поисковых тезаурусов, МБА, отчеты о работе библиотеки, выгрузку данных в различных форматах, поддержку полнотекстовых документов.

Система ДИТ-ИБИС была внедрена в 60-и библиотеках России, среди которых ГЦНМБ, ГПНТБ СО РАН, ЦНСХБ, ГНПБ им. К. Д. Ушинского, РГБИ Минкультуры России, Казанский Республиканский медицинский библиотечно-информационный центр, Библиотека Казанского медицинского университета, Библиотека челябинского медицинского университета, Библиотека Военной медицинской академии г. С-Петербург, 7-ая ЦБС г. Москва и др. На основе ППП ДИТ-ИБИС выпускается CD-ROM с базой данных “Российская медицина”, насчитывающей более 500 тыс. библиографических записей.

В 1995 году компания ДИТ-М приступила к разработке интегрированной информационно-библиотечной системы нового поколения на основе современной мощной СУБД, ориентированной на интернет-технологии. В качестве базового программного обеспечения были выбраны продукты компании Software AG (СУБД ADABAS, Text Reatrival System и др.).

В начале 1998 года компания вышла на рынок информационных технологий с продуктом новой линии под названием OPAC-R. Этот продукт создавался в рамках

программы “Создание общероссийской компьютерной сети библиотек” ЛИБНЕТ. В конце 2000 г. компания создает программный продукт для глобальных электронных библиотек.

Логинов Б.Р., Цветкова И.Б. ©