

УДК 004.9

*Кузьмина Э.В.,
к. п. н., доцент
кафедры математики и информатики
Финансового университета при Правительстве РФ*
*Кухаренко Л.В.,
к.к.н., доцент
кафедры бухгалтерского учета и анализа
Краснодарского филиала РЭУ им. Г.В. Плеханова*

**ФУНКЦИОНАЛЬНЫЕ РЕШЕНИЯ ОРГАНИЗАЦИИ ПОИСКА
ИНФОРМАЦИИ В АВТОМАТИЗИРОВАННЫХ БИБЛИОТЕЧНО-
ИНФОРМАЦИОННЫХ СИСТЕМАХ**

**FUNCTIONAL SOLUTIONS IN ARRANGEMENT OF INFORMATION
SEARCH IN AUTOMATIC LIBRARY-INFORMATION SYSTEMS**

Аннотация: в статье рассматриваются особенности функциональных решений поиска информации в современных автоматизированных библиотечно-информационных системах (АБИС). Предложена группировка решений организации поиска в трех направлениях: анализ и обработке web-коллекций, поиск с использованием специализированных инструментов АБИС – электронного каталога, оцифровка фонда и поиск решений представления и обработки больших данных. Предложено определение информационного поиска для АБИС. Рассмотрено влияние форматов обмена данными на решения по организации поиска информации.

Abstract: The article deals with some features of the functional solutions in arrangement of information search in modern automatic library-information systems (ALIS). It groups the solutions into three directions: analysis and data processing web-collections, search with the helpful special ALIS instruments such as an electronic catalogue, digitizing of fund and finding ways of presentation and large data processing. The definition of information search for ALIS is given. The influence of data exchange formats on solutions of organizing information search is consider as well.

Ключевые слова: Автоматизированная библиотечно-информационная система, АБИС, информационный поиск, электронный каталог, стандарты представления и обмена данными, форматы данных.

Keywords: The automatic library-information system, ALIS, information search, electronic catalog, standards for presentation and data exchange, data formats.

Существование на информационном рынке множества автоматизированных библиотечно-информационных систем (АБИС), обладающих различными качественными и количественными характеристиками, приводит к необходимости оценки их функциональности с точки зрения информационного поиска. Автоматизированные библиотечно-информационные системы это – системы автоматизации библиотечных процессов, которые используются для отслеживания движения библиотечных фондов и справочно-информационного обслуживания читателей [1, С.157].

Одной из основных потребительских функций АБИС является функция поиска информации. Для клиента поиск информации может рассматриваться как совокупность действий, приводящих к выдаче релевантных фактов или документов в ответ на запрос. Можно выделить три направления реализации функциональных решений поиска информации в современных АБИС. Первым направлением является поиск, основанный на анализе и обработке web-коллекций. Вторым направлением выступает поиск с использованием специализированных инструментов АБИС – электронного каталога. Третьим направлением автоматизации библиотек является оцифровка фонда и поиск решений представления и обработки больших данных.

Первое направление обработки web-коллекций АБИС связано со значением термина «поиск информации».

Значение термина «поиск информации» может быть очень широким. Однако для поиска информации может быть использовано следующее определение. Информационный поиск предназначен для поиска неструктурированного материала или документов (обычно текста), который удовлетворяет потребность в информации из больших коллекций, как правило, хранящихся на компьютерах.

Ранее информационный поиск использовался как деятельность, в которой участвовали лишь несколько человек: библиотекари, библиографы, специалисты по научно-технической информации. В современном информационном обществе сотни миллионов людей занимаются поиском информации каждый день, при использовании web-поисковой системы [2, С.299]. Информационный поиск становится преобладающей формой доступа к информации, обгоняя традиционный поиск в реляционных базах данных.

Информационный поиск также может охватывать другие виды данных и информационные проблемы, выходящие за рамки реляционной модели. Термин

«неструктурированные данные» относится к данным, которые не имеют четкой, семантически открытой, легкой для компьютера структуры. Это противоположность структурированных данных, каноническим примером которых является реляционная база данных, применяющаяся обычно для ведения учета товарных запасов и записей о персонале. На самом деле практически нет данных, которые действительно «неструктурированы». Но даже признав, что предполагаемое понятие структуры является открытой структурой, большинство текстов имеют структуру, такую как заголовки, абзацы и сноски, которые обычно представлены в документах с помощью явной разметки. Существует понятие полуструктурированного поиска, такого как поиск документа, в котором заголовок содержит Java, а тело содержит потоки [3, С.58].

Область поиска информации также охватывает поддержку пользователей при просмотре или фильтрации коллекций документов или дальнейшей обработке набора извлеченных документов. При подборе документов используется кластеризация – задача, состоящая в хорошей группировке документов на основе их содержимого [4, С.40]. Это похоже на организацию книг на книжной полке в соответствии с их тематикой. Учитывая набор тем, постоянных информационных потребностей или других категорий (например, пригодность текстов для разных возрастных групп), используется классификация. Классификация - это задача определения того, какой класс, если таковой имеется, принадлежит каждому из множества документов. К данному методу часто обращаются, сначала классифицируя отдельные документы вручную, а затем предполагается выполнение автоматической классификации новых документов на основе заданных признаков.

Информационно-поисковые системы АБИС также можно различать по шкале, в которой они работают. В веб-поиске, система должна обеспечить поиск по миллиардам документов, хранящихся на миллионах компьютеров. Отличительными проблемами такого поиска требуют сбора документов для индексирования, способности создавать системы, работающие эффективно в этих огромных масштабах, обрабатывать конкретные аспекты Интернета, в число которых входят использование гипертекста, и не обманывать провайдеров сайтов, манипулируя содержимым страницы пытаясь повысить их рейтинг в поисковых системах, учитывая коммерческую значимость Интернета. Данные проблемы фокусируются в webcharlink. С другой стороны, существует поиск личной информации [5, С.31]. В последние несколько лет

потребительские операционные системы имеют интегрированный поиск информации (например, Apple Mac OS X Spotlight или мгновенный поиск Windows).

Программы электронной почты обычно не только обеспечивают поиск, но и классифицируют текст: они по меньшей мере предоставляют фильтр спама (нежелательной почты) и обычно также предоставляют ручные или автоматические средства для классификации входящих и исходящих сообщений, чтобы их можно было размещать непосредственно в определенных папках. Отличительные проблемы здесь включают в себя обработку широкого спектра типов документов на типичном персональном компьютере и обеспечение использования бесплатной и достаточно легкой поисковой системы с точки зрения запуска, обработки и использования дискового пространства, так чтобы она могла работать, не перегружая оперативную память машины. Между ними находится пространство корпоративного, институционального и доменного поиска, где может быть предоставлен поиск для таких коллекций, как внутренние документы корпорации, база данных специальных документов или исследовательские статьи. В этом случае документы обычно хранятся в централизованных файловых системах, и одна или несколько выделенных машин будут обеспечивать поиск по коллекции [6, С.35].

Второе направление связано с обеспечением поиска информации по запросам читателей в электронном каталоге. Формирование сводного электронного каталога является характерной чертой автоматизации крупных библиотек. Формирование сводного электронного каталога связано с использованием международных или национальных форматов представления библиографических данных [7, С.142]/ Данное обстоятельство делает подсистему каталогизации в АБИС уникальным автоматизированным решением, что существенно увеличивает стоимость системы. Специалисты ИФЛА разрабатывают новые стандарты объектного представления записей об изданиях в автоматизированных системах, что является попыткой раскрыть содержание как библиотечных, так и музейных фондов.

Современные АИБС должны удовлетворять современным стандартам представления и обмена данными. В автоматизированных библиотечно-библиографических системах используются библиографические форматы.

Формат, по определению, это совокупность элементов данных, их связей и правил их заполнения. Библиографический формат – это такая совокупность взаимосвязанных элементов данных, которая служит для создания и обмена

библиографическими записями в машиночитаемой форме. Принято считать, что правила заполнения элементов данных и определение необходимых связей – это правила составления библиографического описания, международные или национальные, в зависимости от назначения формата [8, С.103].

Однако, все известные существующие форматы действительно опираются на соответствующие правила описания: MARC21 – AACR2, UNIMARC – ISBD, RUSMARC – ISBD и ГОСТ 7.1-84. Однако – и это для подавляющего большинства заинтересованных лиц очевидно, но почему-то мало отражается в практической деятельности – нельзя ставить знак равенства между библиографической карточкой и машиночитаемой библиографической записью, между карточным и электронным каталогом. Все рассматриваемые форматы учитывают это положение, в большей или меньшей степени, и, как результат, содержат ряд элементов, правил, а иногда и связей, не предусмотренных стандартами и правилами составления библиографических описаний [9, С.271].

В настоящее время в международной практике наиболее распространенными форматами данных являются UNIMARC и MARC21 с локальной MARC адаптацией (например, danMARC2, MARC21-Fin, NORMARC). Существуют также некоторые другие локальные форматы, такие как MAB2 (как своего рода «иностранный язык») и OCLC PICA (как своего рода «гибридный язык»), но их уровень использования, что неудивительно, намного ниже, чем у MARC. Наиболее часто используемыми форматами являются UNIMARC и MARC2. UNIMARC основывается на форматах США, MARC поддерживает форматы ЕС. Наиболее часто используемые правила каталогизации являются национальными, за которыми следуют локальные адаптации AACR2. Отдельные учреждения используют модификации или короткие версии национальных кодов каталогизации [10, С.89].

Причины принятия решения о миграции на новые форматы обмена данными достаточно многогранны. Наиболее важной из них является сложность формата MARC21 и тот факт, что многие библиотечные и информационные учреждения хранят записи в формате MARC21 (следовательно, меньше необходимости вносить изменения в загруженные записи). Кроме того, существует поддержка со стороны сообщества MARC21, в том числе полная документация формата, сопоставление с новыми форматами, активные дискуссии и многое другое от поставщиков систем библиотек, имеющих опыт работы с MARC21.

В настоящее время ведущие библиотечные учреждения России ведут тесное сотрудничество с международным библиотечным миром в области форматов библиографических данных. В частности, реализуется интенсивное сотрудничество с ИФЛА - ИФЛА UNIMARC (UP), с Постоянным комитетом UNIMARC (PUC), участие в программе ICABS, с Координационным комитетом по гармонизации MARC (MHCC) и Машиночитаемым библиографическим комитетом информации (MARBI), с целью обмена информацией и приведения европейских идей в международную дискуссию по форматам данных. Из-за важности форматов данных в вопросах каталогизации сотрудничество с органами каталогизации и стандартизации является обязательным, как и сотрудничество с ИФЛА.

Интересны также причины отказа от миграции на новые форматы обмена данными. Некоторые библиотечные учреждения уже перешли на другой формат, довольствуются существующим форматом или используют формат, который очень похож на MARC. Другие ограничены информационной системой и форматом их национального библиотечного ландшафта или систем, поддерживаемых поставщиками библиотечных систем. Еще одним важным соображением по переходу на другие форматы является требование обработки некоторых новых форматов, включая основанные на XML, Dublin Core и ONIX, которые связаны с третьим направлением организации информационного поиска в АБИС.

Третье направление информационного поиска в АБИС: оцифровка фонда и поиск решений представления и обработки больших данных, связаны с исследованиями по использованию реляционной модели данных и концептуальному моделированию.

Разработанные правила каталогизации на основе MARC-формата позволяют использовать реляционную модель данных при создании электронных каталогов и автоматизированных библиотечных систем на основе использования моделей FRBR (Функциональные требования к библиографическим записям).

Описание ресурсов и доступ (RDA) является основной темой в библиотечном мире с начала 2000-х годов. В последние годы к обсуждению добавилась тема функциональных требований к библиографическим записям (FRBR). Один из разработчиков правил RDA Том Делси (Tom Delsey) предложил несколько сценариев реализации этих правил в базах данных. Один из сценариев, в частности, предполагает использование реляционной модели.

Работы по адаптации этих правил в России ведутся, в частности, в национальном информационно-библиотечном центре «ЛИБНЕТ».

Первый проект RDA опубликован в 2003 году. RDA не является заменой MARC; это набор рекомендаций по вводу данных в записи MARC. RDA сама по себе не предназначена для отображения данных в онлайн-каталогах; это стандарт кодирования для каталогизаторов. Если RDA - это набор стандартизованных правил, то FRBR – это концептуальная модель того, как эти правила могут поддерживать расширенные отображения онлайн-каталога.

С 1990-х годов ИФЛА ведет разработку концептуальных моделей для библиографических данных. FRBR. Функциональные требования к библиографическим записям (FRBR) - первая модель, опубликованная в 1998 году. FRBR предоставила «обобщенное» представление о библиографической вселенной, независимо от любого кода или реализации каталогизации. Это позволило более четко понимать библиографические данные, которые не только информировали о разработке кодексов каталогизации, но также способствовали обсуждению и сотрудничеству посредством явного моделирования, которое превосходит локальные различия. Несмотря на то, что FRBR не предназначались для использования в качестве модели данных, они влияли на технологические разработки и реализации с использованием библиографических данных. В 2016 году исследовательской группой ИФЛА по функциональным требованиям для библиографических записей был опубликован заключительный отчет FRBR модели библиографической записи.

Модель FRBR была хорошо воспринята глобальным библиотечным сообществом (FRBR Bibliography). FRBR сосредоточилась, в частности, на библиографических данных, в более узком смысле, и ИФЛА перешла к расширению работы по концептуальному моделированию, включив данные о полномочиях, разработав два расширения для FRBR, одно из которых сосредоточилось на данных о полномочиях, а другое - на данных субъекта полномочий: FRAD (Функциональные требования для данных о полномочиях) и FRSAD (Функциональные требования к данным субъекта).

Эти первые четыре модели иногда известны под общим названием «семейство концептуальных моделей FRBR».

Модель FRBROO начала разрабатываться вскоре после публикации FRBR. Это объектно-ориентированная версия трех исходных моделей сущностей-связей, FRBR,

FRAD и FRSAD. FRBROO лучше подходит для реализации концепций FRBR в объектно-ориентированных инструментах. В нем также содержатся некоторые концепции, которые были неявны в оригинальных моделях, что облегчает переход от концептуальной модели к реальным приложениям.

Модель PRESSOO - формальная онтология, предназначенная для представления библиографической информации о продолжающихся ресурсах, а точнее о сериалах (журналах, газетах и т. п.). PRESSOO стремится предложить ответы на давние проблемы с применением семейства моделей FRBR к сериалам и постоянным ресурсам.

Новейшей концептуальной моделью является ИФЛА LRM, справочная модель библиотеки ИФЛА. IFLA LRM - это концептуальная эталонная модель высокого уровня, разработанная в рамках моделирования сущностей. Это консолидация отдельно разработанных концептуальных моделей: FRBR, FRAD, FRSAD. Она была разработана для устранения несоответствий между тремя отдельными моделями. Результатом является единая, оптимизированная и логически согласованная модель, которая охватывает все аспекты библиографических данных и в то же время обновляет моделирование с помощью современных концептуальных моделей. IFLA LRM была разработана для использования в связанных средах данных, а также для поддержки и популяризации использования библиографических данных в связанных средах данных.

Таким образом, разработанные модели описания библиографических данных позволяют реализовывать автоматизированные библиотечные информационные системы как на основе реляционных, так и объектных систем управления баз данных (СУБД). Модель IFLA LRM позволяет организовать совместимость и передачу данных в связанных средах данных, основанных на интернет-технологиях [11, С.32].

Все направления деятельности библиотек в области автоматизации информационного поиска направлены на сохранность культурного книжного и иного печатного наследия человечества и поиска путей повышения качества обслуживания потенциальных потребителей информации через удовлетворение их информационных потребностей. Небольшие библиотеки зачастую используют некоторые из этих возможностей. Как правило, приобретают один или несколько модулей АБИС, в том числе информационно-поисковые системы, представляющие отдельные модули, связанные с вводом информации о документе и осуществляющие поиск данной информации. При выборе АБИС необходимо учитывать как функционал системы, так и

используемые технологии, включающие форматы библиографических данных, программное обеспечение, включающее тип системы управления базами данных (СУБД), и отзывы клиентов.

Список литературы

1. Кузьмина Э.В., Алексеева С.Н. Эволюция управления библиотечной деятельностью // Кайгородовские чтения. Материалы региональной научно-практической конференции. Краснодарский государственный университет культуры и искусств. – Краснодар, 2010. - С. 156-160.

2. Лобанова А.В., Лазаренко Л.А. Усвоение правил использования предметов в совместной деятельности со взрослым как ключевой путь освоения ребенком культуры. // Проблемы современного педагогического образования. - 2017. № 54-6. - С. 298-304.

3. Салий В.В., Шапошников В.Л., Пьянкова Н.Г., Кузьмина Э.В. Методы анализа больших объемов слабоструктурированной информации. Учебное пособие. – Краснодар, 2017

4. Кузьмина Э.В., Пьянкова Н.Г. Использование технологий хранилищ данных для интеллектуализации фольклорных информационных ресурсов. // Информационные ресурсы России. - 2016. № 6. - С. 39-41.

5. Лобанова А.В., Лазаренко Л.А. О проектных моделях исследования иррационального поведения типичной жертвы экономического мошенничества в условиях перехода к цифромике // Естественно-гуманитарные исследования. - 2017. № 18 (4). - С. 30-33.

6. Николаева И.В. Прикладное программное обеспечение. Учебное пособие / И.В. Николаева. – Краснодар: Эконинвест, 2011. – 123 с.

7. Фролов Р.Н. Развитие взаимодействия с клиентами риэлтерских агентств путем применения web-ориентированных информационных порталов. // В сборнике: Наука и образование в жизни современного общества сборник научных трудов по материалам Международной научно-практической конференции в 14 томах. – Краснодар, 2015. С. 142-143.

8. Салий В.В., Кузьмина Э.В. Применение формализованных методов аналитико-синтетической переработки информации в библиотечно-библиографической

деятельности // Культурная жизнь Юга России. Приложение. - 2015. № 1 (1). - С. 102-105.

9. Кухаренко Л.В. Информационная культура и ее роль в экономической деятельности современного предприятия // Факторы и направления развития туристско-гостиничной индустрии сборник материалов III Всероссийской научно-практической конференции. – Краснодар, 2010. - С. 270-275.

10. Кухаренко Л.В. Сетевая экономика и перспективы развития электронной коммерции в России. // В сборнике: Актуальные проблемы науки Материалы I Всероссийской (заочной) научно-практической конференции (с международным участием). Под общей редакцией А.И. Вострецова. – Воронеж, 2014. - С. 88-92.

11. Горецкая Е.О., Кухаренко Л.В. Новые тенденции интеграционных процессов в России и мировом хозяйстве // Социально-экономическое развитие России: актуальные подходы и перспективные решения Материалы I Международной научно-практической конференции. – Краснодар, 2017. - С. 31-34.