

*Прозоров Иван Евгеньевич* – главный библиограф  
*Кисарова Марина Евгеньевна* – зав. сектором  
Управление библиографическими информационными службами ЦГПБ им. В. В.  
Маяковского (Санкт-Петербург).

Проблемы ведения информационно-поисковых тезаурусов в АБИС: из опыта  
работы библиографов Центральной городской публичной библиотеки  
им. В. В. Маяковского

**[Доклад на 6-й ежегодной международной научно-практической конференции  
«Информационные технологии в профессиональном образовании, научной и библиотечно-  
информационной деятельности», 16 марта 2010 г.,  
Санкт-Петербург, С.-Петерб. гос. университет культуры и искусств]**

В образовательном процессе и профессиональной деятельности большое значение имеет тематический поиск документов. Успешность такого поиска во многом зависит от средств и качества индексирования библиографических записей в информационно-поисковых системах.

Очевидно, что требования к информационному массиву для электронных каталогов и аналитических библиографических баз данных различны. Электронный каталог на практике рассматривается, в первую очередь, как справочно-поисковый аппарат для конкретного фонда документов, выполняющий также учетную функцию. Его информация всегда будет актуальной, т.е. соответствовать отражаемому в нем документальному фонду. Библиографические (и иные документальные) базы данных создаются на определенном документальном фонде (совокупности фондов), однако их целевое назначение предусматривает выборочное отражение документов, соответствующее тематике базы данных, а также включение сведений о документах не связанных с конкретными документальными фондами. Таковы базы данных Российской книжной палаты, отраслевых центров научно-технической информации (ВИНИТИ, ИНИОН, Информкультура, ЦНСХБ, Научной педагогической библиотеки им. К. Д. Ушинского и др.), базы данных (тематические, отдельных видов документов) различных библиотек, информационных служб учреждений, предприятий.

Ведение документальных баз данных (в частности, библиографических) должно предусматривать продолжительность их использования. В условиях динамично меняющихся информационных потребностей пользователей, комплексности целого ряда актуальных научных проблем, изменений структуры научного знания одним из эффективных средств отражения содержания документов и поиска информации являются информационно-поисковые тезаурусы.

Преимущества тезаурусов перед словарями предметных рубрик заключаются в высоких комбинационных возможностях при создании поискового образа документа

любой сложности; простоте использования при формулировке поискового запроса, сравнимой с формулировками естественного языка; большей компактностью массива поискового словаря, что делает отбор пользователем релевантного поискового термина более комфортным. Тогда как адекватная сложная предметная рубрика стремится передать содержание конкретного документа, что ограничивает возможности ее последующего использования при индексировании; требует строгой формализации поискового запроса, и, хоть и обеспечивает высокую точность поиска, но нередко, в ущерб его полноте.

Тезаурус – информационный образ документированного знания, представленного с максимальной степенью аналитичности. Благодаря этому, он становится удобным средством синтеза на уровне поискового предписания. Здесь его комбинационные возможности, если не безграничны, то необычайно широки.

Каждый значимый аспект содержания документа (тематика, имена лиц, названия организаций, хронологический период, географические объекты, события и т.п.) выражается отдельным термином (словом или словосочетанием) и становится самостоятельной точкой доступа. Таким образом, документальный массив базы данных, предполагающий возможности поиска по одному или совокупности семантических признаков, выступает как дружественная поисковая система для пользователей с различным уровнем информационной культуры.

Наибольшее распространение в отечественной практике получили тезаурусы по отдельным научным направлениям, отраслевые, в меньшей степени – межотраслевые тезаурусы. Вместе с тем, полнота представления в тезаурусе специальной лексики и связей между терминами, требования к детальности индексирования (к глубине анализа содержания документов) определяются целевым назначением базы данных, тематикой и видовым составом отражаемых документов и контингентом пользователей. Библиографы Центральной городской публичной библиотеки им. В. В. Маяковского имеют положительный опыт применения политематического тезауруса в библиографической базе данных на протяжении 16 лет.<sup>1</sup>

---

<sup>1</sup> Результаты работы неоднократно освещались в профессиональной печати: Оранская Л. И. Некоторые особенности использования дескрипторного поискового языка в библиографической ИПС универсальной библиотеки // Науч. и техн. б-ки. 1997. № 9. С. 13-23; 1998. № 1. С. 153; Сухарева М. Н. Проблемы отражения краеведческих материалов в библиографической базе данных // Проблемы краеведческой деятельности библиотек / РНБ, Новгород. обл. универс. науч. б-ка. СПб., 2003. С. 107-111; Прозоров И. Е., Кисарова М. Е. Сухарева М. Н. Координатное индексирование и возможности координатного поиска в библиотечных информационно-поисковых системах // Лингвистическое обеспечение информационных ресурсов библиотек, музеев, архивов и других учреждений культуры / ЦГПБ им. В. В. Маяковского. СПб., 2008. С. 95-111 и др.

В отечественной практике создания тезаурусов с 1960-х гг. накоплен серьезный опыт, не получивший, однако, широкого применения в библиотечно-информационной сфере. Отчасти, это связано с недостаточностью внимания разработчиков программного обеспечения библиотечных информационно-поисковых систем к вопросам автоматизации ведения и использования тезаурусов. Многие библиотеки при создании электронных каталогов и библиографических баз данных пошли по пути использования неконтролируемых ключевых слов, «заново открывая проблемы, решение которых описано в литературе 1960-1980-х гг.».<sup>2</sup>

В настоящее время АБИС не обеспечивают полноценного ведения тезауруса в рамках одной АБИС вместе с библиографической базой данных. Удивительно, что в профессиональном сознании информационных специалистов тезаурусы, порой, ассоциируются с объемными и неудобными в использовании томами их печатного варианта. Между тем, эффективность тезаурусов как средства анализа и поиска информации в электронной среде подтверждается практикой их применения в библиографических базах данных ведущих отечественных разработчиков: ИНИОН РАН, Центральной научной сельскохозяйственной библиотеки, Научной педагогической библиотеки им. К. Д. Ушинского, Парламентской библиотеки.<sup>3</sup>

В современной практике ведения тезаурусов в электронной среде в ряде случаев применяется рукописная подготовка словарных статей. Встречаются ситуации, когда база терминов тезауруса не связана, подобно авторитетному файлу, с полями библиографической базы данных, а то и вовсе тезаурус ведется в другом программном обеспечении. Это приводит к ручному вводу дескрипторов (терминов тезауруса) при индексировании библиографических записей, что угрожает точности индексирования.

Политематический тезаурус библиографической базы данных ЦГПБ им. В. В. Маяковского ведется с 1994 г. в программе CDS/ISIS/M (версия 3.0) в соответствии с ГОСТ 7.25 «Тезаурус информационно-поисковый одноязычный» в форме базы данных с возможностью представления терминов в виде текстового файла. Тезаурус состоит из трех лексикографических словарей: алфавитно-семантического, систематического и словаря единичных терминов. Основную часть тезауруса составляет алфавитно-семантический словарь, содержащий в алфавитном порядке дескрипторные статьи и термины-недескрипторы (аскрипторы) от которых сделаны ссылки.

---

<sup>2</sup> Лаврёнова О. А. Методика разработки информационно-поискового тезауруса // Библиотековедение. 2000. № 6. С. 39-42. Более развернуто методика разработки тезауруса представлена автором в одноименной книге (М. : Пашков дом, 2001. 54 с.).

<sup>3</sup> Москаленко Т. А, Мякова Н. А. Информационно-поисковый тезаурус Парламентской библиотеки : этапы разработки, ведение, применение и дальнейшие перспективы // Науч. и техн. б-ки. 2009. № 3. С. 18-22.

Каждому дескриптору присвоены категориальный и тематический индексы. В состав дескрипторной статьи входят следующие категории лексических единиц: заглавный дескриптор, перечень аскрипторов (недескрипторов) – синонимов и условных синонимов, перечень вышестоящих дескрипторов (родовых, т. е. более широких по объему понятий), перечень нижестоящих (видовых) дескрипторов, обратные ссылки от аскрипторов, значение которых передается комбинацией дескрипторов или одним из альтернативных дескрипторов. В дескрипторную статью нередко включается определение заглавного дескриптора (с указанием источника определения в виде цифровой отсылки к списку использованных изданий и сайтов), что способствует однозначному пониманию термина при индексировании документов.

Для уточнения смысла терминов также используются реляторы, поясняющие значение (ДОГОНЯЮЩЕЕ РАЗВИТИЕ, соц.-экон.) или указывающие область знания, к которой принадлежит термин (ДОКУМЕНТЫ [фин.], ДОКУМЕНТЫ [полит.], ЧЕРНЫЕ ДЫРЫ [астр.]).

Пример дескрипторной статьи «Валютная политика» из базы данных «Тезаурус»

(по полям):

020		0758
007		Ф0330
012		41
001		ВАЛЮТНАЯ ПОЛИТИКА
009		сост. часть экономич. политики гос-ва или междунар. фин. организаций - воздействие на валюту, валютные курсы, валютные операции; осн. инструменты - в. интервенция, в. ограничения, в. резервы, в. субсидирование, в. паритеты [2]
002	с	Валютное регулирование
002	с	Валютные отношения
002	с	Валютный коридор
003	в	ФИНАНСОВАЯ ПОЛИТИКА
003	в	ЭКОНОМИЧЕСКАЯ ПОЛИТИКА
004	н	АНТИИНФЛЯЦИОННЫЕ МЕРЫ
004	н	ДЕНЕЖНЫЕ РЕФОРМЫ
005	а	ВАЛЮТА
005	а	ВАЛЮТНЫЕ КУРСЫ
005	а	ВАЛЮТНЫЕ СИСТЕМЫ
005	а	ГОСУДАРСТВЕННОЕ РЕГУЛИРОВАНИЕ [экон.]
005	а	ДЕНЕЖНО-КРЕДИТНАЯ ПОЛИТИКА
005	а	ИНОСТРАННАЯ ВАЛЮТА
006	сК	Валютный контроль
015	сА	Денежная политика
021		000679
072		1999-07-01

(В представленном примере поле 020 – номер записи, 001 – заглавный термин (заголовок авторитетной записи); 002 – синоним-аскриптор; 003 – вышестоящий

*дескриптор; 004 – нижестоящий дескриптор; 005 – ассоциативный дескриптор; 006 – обратная ссылка к понятию, значение которого передается комбинацией заглавного дескриптора данной статьи с другим дескриптором; 007 – тематический индекс; 012 – категориальный индекс; 009 – примечания; 015 – обратная ссылка к понятию, значение которого передается заглавным дескриптором или иным термином из числа альтернативных; 021 – идентификатор записи; 072 – дата ввода).*

Пример дескрипторной статьи «Валютная политика» из базы данных «Тезаурус»(текстовый справочный файл):

~ ВАЛЮТНАЯ ПОЛИТИКА

(сост. часть экономич. политики гос-ва или междунар. фин. организаций - воздействие на валюту, валютные курсы, валютные операции; осн. инструменты - в. интервенция, в. ограничения, в. резервы, в. субсидирование, в. паритеты [2])

с Валютное регулирование

Валютные отношения

Валютный коридор

в ФИНАНСОВАЯ ПОЛИТИКА

ЭКОНОМИЧЕСКАЯ ПОЛИТИКА

н АНТИИНФЛЯЦИОННЫЕ МЕРЫ

ДЕНЕЖНЫЕ РЕФОРМЫ

а ВАЛЮТА

ВАЛЮТНЫЕ КУРСЫ

ВАЛЮТНЫЕ СИСТЕМЫ

ГОСУДАРСТВЕННОЕ РЕГУЛИРОВАНИЕ [экон.]

ДЕНЕЖНО-КРЕДИТНАЯ ПОЛИТИКА

ИНОСТРАННАЯ ВАЛЮТА

сК Валютный контроль

сА Денежная политика

В самостоятельный алфавитно-семантический словарь изначально были выведены так называемые единичные термины (имена собственные: названия исторических событий, географические наименования, имена лиц, названия организаций и т.п.). Для этой категории терминов свойственны более высокие темпы роста поискового словаря. Кроме того, они являются терминами последнего уровня в иерархии понятий – в информационно-поисковом словаре они связаны только с вышестоящими дескрипторами, обозначающими более широкую предметную область. В библиографической базе данных единичные термины реализованы в виде предметных рубрик (ПР) (ПР-имя лица, ПР-коллектив, географическая ПР, тематическая ПР). Ввод ПР в состав поискового образа документа, как правило, сопровождается включением адекватного содержанию статьи дескриптора из числа вышестоящих по отношению к взятой ПР. С 2001 г. ЦГПБ им. В. В.

Маяковского перешла на АБИС ИРБИС ГПНТБ России. По мере развития программы в ней появились средства ведения контролируемых поисковых словарей в форме баз данных авторитетных файлов. Использование авторитетных файлов в АБИС позволяет обращаться к терминам поисковых словарей непосредственно из соответствующего поля рабочего листа ввода данных библиографической базы данных. Это ускоряет процесс индексирования и избавляет библиографов от клавиатурного ввода терминов, а базу данных от ошибок и опечаток.

В настоящее время в ИРБИС уже используются при индексировании библиографической БД авторитетный файл (АФ) имен лиц (с декабря 2006 г.) и АФ географических рубрик (с нач. 2009 г.). Ведется работа по переводу терминов тематических ПР и наименований коллективов из прежней программы в базы данных соответствующих авторитетных файлов в АБИС ИРБИС. Программное обеспечение позволяет полностью автоматизировать процессы ведения авторитетных файлов этой категории терминов: включение новых терминов, автоматический контроль на дублетность, редактирование, использование при индексировании.

Вместе с тем, остается нерешенным вопрос о ведении тезауруса в рамках АБИС ИРБИС. Программой предусмотрено соответствующее поле для дескрипторов («965: Дескрипторы»). Однако в базе данных тезауруса (функционально является авторитетным файлом) до сих пор не решены проблемы полноценного представления дескрипторной статьи. Для вышестоящих дескрипторов (их в дескрипторной статье может быть несколько) предусмотрено неповторяющееся поле для ввода лишь одного термина (в тезаурусе, например, к дескриптору ВЕДЫ даны два вышестоящих дескриптора ЛИТЕРАТУРНЫЕ ПАМЯТНИКИ и РЕЛИГИОЗНЫЕ ТЕКСТЫ, к дескриптору СОБАКИ – три: ДОМАШНИЕ ЖИВОТНЫЕ, МЛЕКОПИТАЮЩИЕ, ХИЩНИКИ, и таких примеров немало). Не предусмотрены поля для нижестоящих дескрипторов, а их в дескрипторной статье может быть до нескольких десятков (так, к дескриптору НАРОДЫ даны 84 нижестоящих дескриптора). Отсутствуют поля для синонимов-аскрипторов (отражение этих терминов в дескрипторной статье позволяет библиографу выбрать наиболее адекватный содержанию статьи термин), нет полей для систематических и категориальных индексов и др. Не предусмотрена возможность автоматической простановки обратных парадигматических связей, создания систематического указателя и т.д.

Положительным следует считать наличие в базе данных тезауруса в АБИС ИРБИС повторяющегося поля для ассоциативных дескрипторов (связанных с заглавным термином по принципу «причина-следствие», «процесс-субъект», «процесс-объект»,

«функционального сходства» и др.). Их число также может достигать нескольких десятков.

Пример дескрипторной статьи в АБИС ИРБИС:

The screenshot shows the IRBIS database interface. The title bar reads "ИРБИС - АРМ 'Каталогизатор' - Центральная городская публичная библиотека им. В.В. Маяковского, г. Са...". The main window displays a search result for "ВАЛЮТНАЯ ПОЛИТИКА" (Terms fully) - 1. The interface includes a menu bar (База данных, Корректировка, Поиск, Просмотр, Сервис, Помощь), a toolbar, and a search bar. Below the search bar, there are dropdown menus for "Термины полностью", "Оптимизированный", and "ТЕЗ - Основной РЛ".

Название элемента	№	Значение
<b>Основной термин</b>		<b>ВАЛЮТНАЯ ПОЛИТИКА</b>
Вид - Дескриптор(1)/Недескриптор(0)	1	
Вышестоящий дескриптор		ФИНАНСОВАЯ ПОЛИТИКА
Ассоциативные дескрипторы	1	ВАЛЮТА
	2	ВАЛЮТНЫЕ КУРСЫ
	3	ВАЛЮТНЫЕ СИСТЕМЫ
	4	ГОСУДАРСТВЕННОЕ РЕГУЛИРОВАНИЕ [экон.]
	5	ДЕНЕЖНО-КРЕДИТНАЯ ПОЛИТИКА

Below the table, there are tabs for "Основной" and "Добавочные". The "Основной" tab is active, showing a detailed description of the term "ВАЛЮТНАЯ ПОЛИТИКА - Дескриптор". The description includes the level (5), the parent term (ФИНАНСОВАЯ ПОЛИТИКА), and a list of associative terms: ВАЛЮТА, ВАЛЮТНЫЕ КУРСЫ, ВАЛЮТНЫЕ СИСТЕМЫ, ГОСУДАРСТВЕННОЕ РЕГУЛИРОВАНИЕ [экон.], ДЕНЕЖНО-КРЕДИТНАЯ ПОЛИТИКА, ИНОСТРАННАЯ ВАЛЮТА. A note explains that it is a part of economic policy of a state or international financial organizations, affecting exchange rates, currency operations, etc.

At the bottom, there is a status bar with fields for "БД: ТЕЗ Макс.MFN: 8141", "Текущий MFN: 843", "Отмечено - 0", and a timestamp "19:53 00:07".

Дескрипторный словарь к аналитической библиографической базе данных ЦГПБ им. В. В. Маяковского в усеченном виде был введен в базу данных тезауруса в АБИС ИРБИС в конце 2006 г. стараниями предыдущего менеджера библиографической базы данных А. А. Певкур и используется в настоящее время для автоматизированного индексирования библиографических записей (заполнения поля 965), что существенно ускорило процесс аналитической обработки изданий. Для этих целей было использовано предлагаемое АБИС ИРБИС «дерево терминов» базы данных тезауруса, в котором было создано более 70 иерархических рядов терминов. Предлагаемый классификационный принцип деления терминов не соответствовал принципам построения тезауруса. Многие термины были отнесены к тому или иному классу понятий условно. Сегодня при индексировании библиограф может найти необходимый термин как методом словарного поиска в тезаурусе, так и через «дерево терминов». Но наиболее точно и адекватно дескриптор может быть найден в текстовом файле тезауруса, который используется

библиографами в качестве справочника. В текстовом файле в MS Word алфавитно-семантический словарь терминов представлен полностью со всеми связями.

Ведение тезауруса продолжается в программе ISIS с использованием пакета прикладных программ и текстового редактора: автоматическое сравнение словарей библиографической и авторитетной БД и выявление новых терминов, создание и редактирование дескрипторных статей, проверка на дублетность, создание систематического указателя, вывод терминов словаря в текстовый файл. Кроме того, автоматически проставляются необходимые ссылки (в том числе вновь введенный термин) в тех дескрипторных статьях, заглавные термины которых введены в состав новой дескрипторной статьи. Это существенно облегчает словарную работу и исключает ошибки. Новые термины и изменения в дескрипторных статьях передаются редактором тезауруса менеджеру библиографической базы и вручную вводятся в состав тезауруса в АБИС ИРБИС.

Реализация полноценного механизма автоматизированного ведения тезауруса в рамках АБИС ИРБИС, по нашему мнению, создаст ей дополнительное конкурентное преимущество и расширит круг потенциальных пользователей программы, а также сможет послужить моделью для аналогичных решений в других отечественных АБИС.

Возможности применения тезаурусов не ограничены только библиографическими базами данных. Они с успехом могут заменить сложные предметные рубрики, с одной стороны, и неупорядоченный массив ключевых слов, с другой, в качестве основного поискового средства в электронных каталогах библиотек. Все зависит от того, насколько создатели электронных каталогов будут заинтересованы в развитии их поискового потенциала (а не только в обеспечении учета документов и организации фонда), в полноценном и точном раскрытии библиотечного фонда с учетом характера информационных запросов основных категорий своих пользователей. Необходимо помнить, что пользователь нуждается не в поиске документов, а в нахождении необходимой информации, что требует многоаспектного раскрытия содержания документов, полноценным средством которого выступают тезаурусы.

Противоречие между уникальными возможностями тезаурусов как поискового средства и возможностями программного обеспечения, безусловно, явление временное. Убеждены, что автоматизация всех процессов ведения тезауруса в АБИС относится к перспективному направлению их развития.