

Основные проектные решения создания национальной базы данных авторитетных записей

С 4 кв. 2001 г. Национальной библиотекой Беларуси с привлечением ведущих специалистов других организаций выполнялся проект по разработке автоматизированной системы создания национальных БД авторитетных записей. Проект разрабатывался в рамках Государственной программы информатизации «Электронная Беларусь» (задание 89). В декабре 2005 г. работы завершены. Публикуем краткие итоги работы.

Проект выполнялся в целях:

– разработки опытного образца автоматизированной информационной системы создания и ведения национальных баз данных авторитетных/нормативных записей; которая должна обеспечивать технологические процессы формирования, ведения и использования записей этого типа;

– создания опытного образца национальных баз данных авторитетных/нормативных записей на основные типы объектов (энтитетов), отражаемых в поисковых полях документографических БД.

Нелишне напомнить, что БД авторитетных записей обеспечивают более эффективный поиск в электронных каталогах библиотек, т.е. в тех документографических БД, которые, как правило, не ограничены ни тематикой, ни видовым охватом документопотока. Чем больше объем электронного каталога, тем заметнее эффект применения авторитетных записей.

Разрабатываемая в рамках проекта технология рассчитана на кооперацию учреждений, участвующих в генерации национальных интегрированных документографических ресурсов. БД авторитетных записей является одной из важнейших составляющих создаваемого в Беларуси сводного электронного каталога библиотек, который предназначен для свободного обслуживания через Интернет пользователей любых категорий и сфер занятости.

Результаты проекта будут использоваться библиотеками всех уровней Республики Беларусь и прежде всего – участницами сводного электронного каталога; они могут быть использованы также различными ведомствами, архивами и музеями при создании документографических электронных ресурсов.

Программное обеспечение для хранения и использования БД авторитетных записей создано в среде СУБД Oracle 9i, а **программный комплекс** по формированию, редактированию и загрузке авторитетных записей рассчитан на его применение любым участником, который имеет выход в Интернет и располагает стандартным браузером. Поэтому не существует никаких технических препятствий для использования любой библиотекой Беларуси программного комплекса по формированию и редактированию ав-

торитетных записей. Основные отличительные особенности программного обеспечения, созданного Национальным центром информационных ресурсов и технологий НАН Беларуси, следующие:

1) невысокие требования к конфигурации ПТК для прикладной программы создания и ведения авторитетных записей MARCEDIT: Windows' 98 и выше, Internet Explorer 5.0 и выше, Microsoft Office, если подключается проверка орфографии;

2) два режима пополнения БД авторитетных записей: on-line и пакетный;

3) загрузка готовых записей непосредственно в БД или накопление их в файле;

4) создание записей вручную или редактирование ранее подготовленных записей путем выгрузки их из БД в макет ввода;

5) обращение к нескольким БД при создании новой записи;

6) автоматическая установка обратных смысловых связей между заголовками;

7) модификация пользователем интерфейса поиска (в пределах индексных полей БД);

8) ведение БД авторитетных записей в условиях работающего каталога (реализована автоматическая корректировка ссылок в библиографических записях на авторитетные при замене последних).

Национальные БД авторитетных записей включают авторитетные, ссылочные и справочные записи в формате BELMARC/A, которые созданы белорусскими учреждениями независимо от того, отражают эти записи национальные объекты или нет.

Наиболее важные **особенности национальных БД авторитетных записей** можно свести к следующим характеристикам:

1) структура записей соответствует формату UNIMARC/A в версии 2001 г., а это означает, что БД включают расширенный репертуар объектов и записи содержат больше данных;

2) представлены все типы объектов, предусмотренные форматом UNIMARC/A;

3) авторитетные записи могут подготавливаться на русском и белорусском языках, что позволяет вести каталогизацию и обслуживание пользователей на обоих государственных языках, при этом документ каталогизируется только на одном из языков;

4) заголовки в составе записей могут быть представлены на любом языке с использованием кодировки UNICODE;

5) включаются авторитетные записи как на национальные, так и на зарубежные объекты, но полные записи создаются вручную в основном на национальные объекты; для описания зарубежных объектов записи предполагается заимствовать;

Таблица 1

Тип объекта	Авторитетный/нормативный заголовок
имя лица	200 #0\$aСофья Слуцкая\$f1585-1612
коллектив (постоянный или временный)	210 02\$aИнститут математики \$cМинск
торговая марка (товарный знак)	216 ##\$7ca0yba0e \$aHORIZONT \$f1976- \$cтоварный знак
адм.-территориальное или географическое название	215 ##\$aЧервень, город (Минская область) 215 ##\$aДняпро, рака
родовое имя	220 ##\$aРадзівілы, род\$f15 ст. –
унифицированное заглавие документа	230 ##\$a«Паўлінка»\$lспектакль \$nПаст. Гродзенскі абласны драматычны тэатр \$q1979
имя/заглавие	240 ##\$1 200#1\$aСеменяко \$bЮ.В.\$gЮрий Владимирович \$1 230 ## \$aПавлінка \$lмузыкальная комедия
имя /обобщающее унифицированное заглавие	245 ##\$1 200#1\$aЧехов \$bА.П.\$gАнтон Павлович \$1 235 #1 \$a Пьесы
тематический предмет	250 ##\$aСТАБИЛИЗАТОРЫ (электротехн.)
место издания как точка доступа	260 ##\$dМінск
форма, жанр, физические характеристики документа	280 ##\$aЛИРИЧЕСКИЕ СТИХОТВОРЕНИЯ

б) вывод записей пользователю для просмотра возможен в трех формах: в виде списка, справки, в формате BELMARC/A.

Все типы имен и образцы заголовков национальных БД авторитетных записей представлены в таблице 1. На сером фоне приведены типы объектов, впервые появившиеся в версии формата UNIMARC/A 2001 г.

Все записи национальных авторитетных БД должны быть получены в результате интеллектуальной работы по формированию либо редактированию самих записей или, как минимум, их главных, исходных заголовков. Объем работ по созданию национальных БД авторитетных записей настолько велик, что, наряду с ручным формированием авторитетных записей, нельзя исключить технологическую возможность их программного формирования на основе правильно структурированного заголовка, который каталогизатор помещает в поле точки доступа библиографической записи. Программное обеспечение такую возможность дает.

Для всех объектов, кроме тематических терминов (тематических предметов), в национальной базе авторитетных данных применяется единая система заголовков, которой пользуются все участники корпорации. Для обозначения тематических предметов в системе допускается параллельное использование разными участниками нескольких принимаемых корпорацией систем предметизации, т.е. нормативных словарей, тезаурусов или перечней предметных рубрик. Это означает, что тематический поисковый образ одного документа в СЭК может иметь несколько версий в соответствии с разными системами индексирования, если запись обрабатывали и дополняли несколько участников корпорации. Пользователь будет выбирать тот поисковый образ документа, ко-

торый ему больше подходит. Участники корпорации смогут заимствовать термины друг у друга, а также связывать термины разных словарей.

Предполагаемая организация СЭК такова, что поиск в нем может обеспечиваться не только записями национальных авторитетных БД, но и так называемыми прототипами авторитетных записей, которые получены алгоритмически и не подвергались интеллектуальной обработке.

Ниже на рис. 1 приводится фрагмент списка заголовков массива прототипов авторитетных записей, полученного на основе машинного словаря персональных ЭК НББ. Хочу заметить, что если взять машинный словарь любой другой библиотеки при условии, что он не контролировался строго, то картина будет не лучше.

Прототипы авторитетных записей создаются программно на основе наполнения поисковых полей библиографических записей библиотечных электронных каталогов прошлых лет. Эти поля формировались в условиях недостаточного нормативного контроля поисковых полей или его полного отсутствия. Поэтому один и тот же объект в массиве прототипов авторитетных записей может быть представлен несколькими формами заголовков, не сводимых программно к одной форме. На каждую форму одного и того же заголовка алгоритмически строится отдельный прототип авторитетной записи. Все прототипы включают минимальный обязательный набор полей, требуемый форматом BELMARC/A.

Массив прототипов авторитетных записей может послужить также дополнительным источником при формировании вручную новых авторитетных записей, если каталогизатор сталкивается с полным отсутствием сведений об объекте в каталогизируемом источнике и не может дифференцировать объект.

Состав комплекса БД. Создание точек доступа библиографических записей технологически должен обеспечивать комплекс из двух корпоративных авторитетных БД, в который входят:

1) рабочая база авторитетных данных, ее записи связаны с полями точек доступа библиографических записей;

2) база авторитетных данных, выполняющая функцию **накопителя** заранее подготовленных авторитетных записей. Эта вторая БД является источником пополнения рабочей базы, ее записи соответствуют требованиям формата BELMARC/A, но не связаны с библиографическими записями корпоративного ресурса.

Обе БД – и рабочая, и накопительная – включают только те записи, которые прошли интеллектуальную обработку или получены на основе хорошо структурированных записей фактографических БД (ФБД).

На рис. 2 приводится фрагмент списка заголовков авторитетных записей БД-накопителя. Представленный фрагмент получен алгоритмически на основе ФБД «Беларусь у падзеях і асобах», созданной НББ в предыдущие годы. Хорошо видна разница в форме заголовков на Якуба Коласа в этом списке и в списке заголовков для этого же писателя из массива прототипов авторитетных записей (рис. 1).

Рабочая база авторитетных данных формируется в процессе каталогизации документов и обеспечивает создание контролируемых поисковых полей библиографических записей.

БД-накопитель пополняется несколькими разными способами:

1) конвертирование в структуру авторитетных данных тех фактографических БД, которые имеются у библиотек и содержат необходимые сведения;

2) плановая подготовка библиотеками массивов авторитетных записей по какой-либо предметной области (например, записи на имена ученых какого-либо вуза или НИИ);

3) заимствование авторитетных записей у зарубежных центров.

Способ формирования накопительной БД авторитетных записей делает возможным наличие в ней нескольких записей на один и тот же объект, которые будут включать

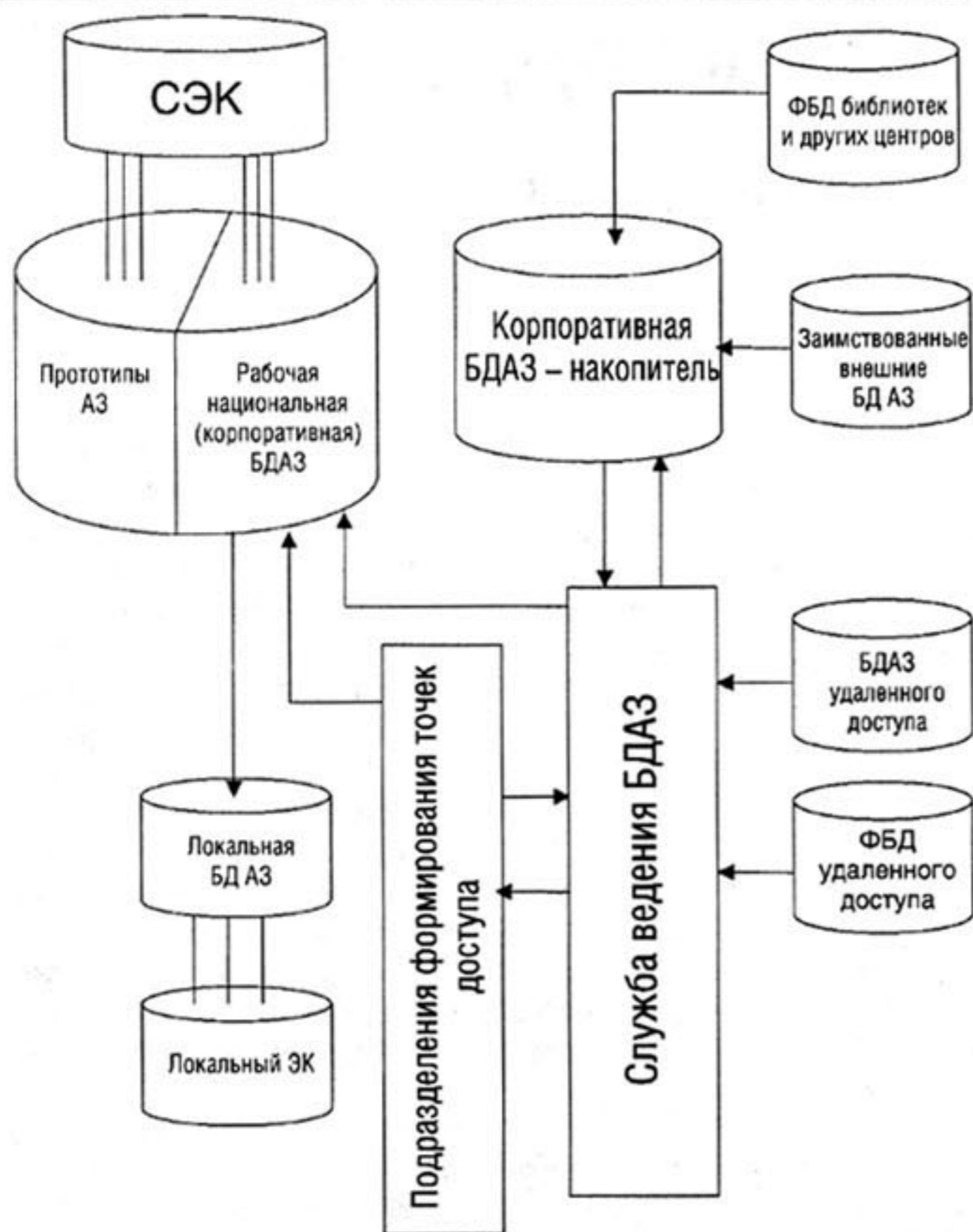
Рис. 1. Фрагмент списка заголовков массива прототипов авторитетных записей

	MARC	Колас
	●	Колас, Віктар см. Колас, Віктор
	●	Колас, Людміла см. Колас, Людміла
	●	Колас, Людміла Якаўлеўна см. Колас, Людміла Яковлевна
I	MARC	Колас, Людміла
I	MARC	Колас, Людміла Яковлевна
	●	Колас, М.Я. см. Колас, М.Я.
I	MARC	КОЛАС (МІЦКЕВІЧ КАНСТАНЦІН М ХАЙЛАВІЧ), ЯКУБ
I	MARC	Колас (Міцкевіч Канстанцін Міхайлавіч), Якуб
I	MARC	Колас (Міцкевіч Канстанцін Міхайлавіч;18821956), Якуб
	MARC	КОЛАС (МІЦКЕВІЧ КАНСТАНЦІН МІХАЙЛАВІЧ;18821956), ЯКУБ
I	MARC	КОЛАС (МІЦКЕВІЧ КАНСТАНЦІН), ЯКУБ
	MARC	КОЛАС (МИЦКЕВИЧ КОНСТАНТИН МИХАЙЛОВИЧ; 182219), ЯКУБ
	MARC	КОЛАС (МИЦКЕВИЧ КОНСТАНТИН МИХАЙЛОВИЧ;18821956), ЯКУБ
	
	MARC	Купала
	MARC	Купала (Луцэвіч Іван Дамінікавіч), Янка
	MARC	Купала (Луцэвіч Іван Домінікавіч)
I	MARC	Купала (Луцэвіч Іван Домінікавіч), Янка
	MARC	КУПАЛА (ЛУЦЕВИЧ ИВАН ДОМИНИКОВИЧ;18821942), ЯНКА
I	MARC	Купала (Луцэвіч Іван Дамінік), Янка
I	MARC	Купала (Луцэвіч Іван Дамінікавіч), Янка
	MARC	Купала (Луцэвіч Іван Дамінікавіч;18821942), Янка
I	MARC	Купала (Луцэвіч Іван Дамінікавіч), Янка
	●	Купала, Я. см. Купала, Янка
I	MARC	Купала, Янка
	●	Купала, Янка см. Купала (Луцэвіч Іван Домінікавіч), Янка
	●	Купала, Янка см. Купала (Луцэвіч Іван Дамінік), Янка

Рис. 2. Фрагмент списка заголовков БД-накопителя

I	MARC	Колас, Іван Андрэвіч (адзін з арганізатараў партызанскага руху на Палессі ў Вялікую Айчынную вайну; 1923)
I	MARC	Колас, Андрэй Данілавіч (журналіст; 1918)
I	MARC	Колас, Валеры Паўлавіч (вучоны-фізік; 1950)
I	MARC	Колас, Георгій Данілавіч (літаратурны і тэатральны крытык; 1929/1994)
I	MARC	Колас, Людміла Якаўлеўна (спявачка; 1950)
I	MARC	Колас, Міхаіл Яўхімавіч (акцёр і рэжысёр; 1935)
I	MARC	Колас, Якуб (паэт; 18821956)
I	MARC	Коласаў, Станіслаў Васілевіч (вучоны ў галіне радыёфізікі; 1946)
I	MARC	Колб, Уладзімір Гаўрылавіч (вучоны ў галіне клінічнай біяхіміі; 1929)
	●	Коленніков, Васілій Фёдаравіч см. Коленнікаў, Васіль Фёдаравіч (удзельнік вызвалення Гомеля; 1921)
I	MARC	Колесаў, Уладзімір Пятровіч (рэжысёр; 1939)
I	MARC	Колесава, Алена Фёдарэўна (удзельніца партызанскага руху ў Мінскай вобласці ў Вялікую Айчынную вайну; 1920/1942)
I	MARC	Колесень, Віктар Пятровіч (вучоны ў галіне жыллагадоўлі; 1944)
I	MARC	Колеснік, Владимир Андрэевіч см. Колеснік, Уладзімір Андрэевіч (крытык; 1922/1994)
	●	Колеснік, Іван Іванавіч см. Колеснік, Іван Іванавіч (паэт; 1932/1979)
	●	Колесніков, Віталь Міхайлавіч см. Колеснікаў, Віталь Міхайлавіч (ваенны дзеяч; 1922/1964)
	●	Колмаков, Юрый Павлавіч см. Колмакоў, Юрый Паўлавіч (спартсмен; 1945)
	●	Колоколов, Юрий Борисович см. Колакалаў, Юрый Барысавіч (партыйны і дзяржаўны дзеяч Беларусі; 1926)
	●	Коломійцава, Татяна Міхайловна см. Каламійцава, Татяна Міхайлаўна (дыржор; 1914/1994)
I	MARC	Колікаў, Мікалай Васілевіч (расійскі і беларускі вучоны ў галіне медыцынскіх і ветэрынарных навук; 1888/1964)

Рис. 3. Роль и соотношение БД различного уровня и назначения в системе авторитетного контроля и формирования национальных БД авторитетных/нормативных записей



эквивалентные по содержанию установленные заголовки. Поэтому каталогизатор, заимствуя данные из накопительной БД, может выбрать более полную запись из нескольких имеющихся или дополнить одну выбранную запись данными из других записей.

Обе корпоративные БД авторитетных записей – рабочая и накопительная – должны быть доступны через Интернет.

Все авторитетные записи, непосредственно связанные с библиографическими записями и обеспечивающие поиск в сводном электронном каталоге, – записи рабочей корпоративной БД и прототипы авторитетных записей – физически хранятся в одной БД, а логически выделяются с помощью фильтра на основе таких признаков, как статус записи, статус исходного заголовка и дата создания записи.

Рабочая и накопительная БД авторитетных записей – это и физически разные БД (рис. 3).

Порядок формирования национальных БД авторитетных записей

Представляется, что основным принципом в организации создания национальных БД авторитетных записей следует принять следующий: при формировании точек доступа библиографических записей все участники корпорации должны использовать только те авторитетные записи, которые уже загружены в рабочую БД и которым присвоен идентификатор национальной рабочей авторитетной БД.

Это означает, что каждая из локальных автори-

тетных БД библиотек, входящих в корпорацию, должна представлять собой копию некой составной части рабочей корпоративной БД, именно той её части, которая отражает точки доступа библиографических записей их локальных электронных каталогов.

Основной объем работ по формированию национальных баз авторитетных записей должны выполнять подразделения каталогизации и службы ведения авторитетных БД тех библиотек республики, которые будут входить в корпорацию. При этом они должны постоянно взаимодействовать друг с другом как опосредованно, через создаваемый ими совместно информационный продукт, так и непосредственно, совместно решая методические и производственные вопросы.

Корпоративное взаимодействие библиотек республики будет организовывать Центр корпоративной каталогизации, который предполагается сформировать при Национальной библиотеке Беларуси.

В библиотеках функции по формированию авторитетных записей должны быть распределены между подразделением, формирующим поля точек доступа в библиографических записях, и службой ведения авторитетной БД. При этом каталогизаторы, оформляющие точки доступа, используют готовые записи или, в случае их отсутствия, предлагают новые авторитетные записи с заголовком статуса «предварительный», а библиографы службы ведения редактируют запись и переводят заголовок, если объект дифференцирован, в разряд установленных, принятых заголовков. Каталогизаторы, предлагающие новую запись, должны извлечь всю доступную информацию из обрабатываемого документа, а библиограф службы ведения авторитетной БД дополняет запись сведениями, пользуясь всеми вспомогательными источниками, включая корпоративную накопительную базу авторитетных записей и другие доступные ему БД.

В результате выполнения проекта библиотеки республики будут обеспечены всем необходимым инструментарием создания национальных БД авторитетных записей: как программными средствами, так и технологической документацией. Кроме того, на протяжении более полутора лет в Национальной библиотеке Беларуси ведется практическая апробация проектных решений и всех обеспечивающих средств автоматизированной информационной системы создания и ведения национальных баз данных авторитетных/нормативных записей силами специально созданного структурного подразделения.

Summary

The article dwells on creation of authority files in the National library of Belarus. It illustrates examples from the authority databases and discusses the processing technology.