



Александр Михальченко

ПРИНЦИПЫ СОЗДАНИЯ НАУЧНО-СПРАВОЧНОГО АППАРАТА К НЕСВИЖСКОМУ АРХИВУ

1. Целесообразность и практическая необходимость создания научно-справочного аппарата к Несвижскому архиву на основе использования информационных технологий

Для современного этапа общественного развития характерны, с одной стороны, все возрастающий ритм деятельности человеческого сообщества в целом и, с другой стороны, все возрастающие объемы информации, которые надо обработать конкретному человеку для успешного существования в этом постоянно ускоряющемся обществе. Это ускорение в полной мере коснулось и архивной отрасли.

Можно отметить неуклонно растущий интерес к ведомственным и государственным архивам, содержащим уникальные запасы исторической и справочной информации. Для того, чтобы облегчить поиск нужных сведений в этом безбрежном информационном мире, особенно исследователю, который не является архивистом-профессионалом, и создается научно-справочный аппарат (НСА).

Следует признать, что существующая в архивах практика работы с традиционными справочниками (описями, каталогами, путеводителями, указателями и др.) уже перестала обеспечивать удовлетворительную оперативность, полноту и достоверность выполнения запросов к содержащим значительное количество документов фондам. Особенно это касается архивов, обладающих информацией, которая имеет международное значение, в случаях, когда потенциальные потребители информации находятся на значительном удалении от архивов.

Выходом из такого положения является использование новых информационных технологий, позволяющих создавать электронные архивы документов. Следует отметить, что создание полноценного электронного архива, включающего полные тексты документов (или их оцифрованные обзоры) требует весьма значительных финансовых и временных затрат, что часто,

по различным причинам, является мало приемлемым. Одним из путей решения этой проблемы может послужить совмещение традиционных и новых технологий обработки архивных документов. Так, в ходе проведения традиционной экспертизы документов (дел) могут создаваться их формализованные описания и на их основе строится автоматизированная поисковая система, позволяющая в короткие сроки найти нужную информацию и выйти на конкретный документ (дело). Фактически создается система электронного НСА, который путем распространения на машинных носителях информации (например, на компакт-дисках) или через информационные системы непосредственного доступа (например, Интернет) становится доступен широкому кругу территориально удаленных от архивов пользователей. Подобные системы могут быть созданы значительно быстрее, чем системы, содержащие полные тексты документов, к тому же стоимость их окажется значительно ниже.

В СНГ уже имеется опыт работы над подобным международным проектом по созданию в электронном виде системы НСА к международному архиву Коминтерна.

Специалисты Национального исторического архива Беларуси с любезного разрешения руководства Российского государственного архива социально-политической истории (РГАСПИ) ознакомились с методикой описания документов Коминтерна и программным обеспечением проекта. Нужно отметить огромную работу, проделанную специалистами этого архива по созданию методики дифференцированного описания дел и документов, доработке программы и пр. Это является хорошей отправной точкой при разработке подобных проектов. В итоге нам стали яснее видны отличия в составе, содержании фондов и, следовательно, в методике описания дел.

Уже при предварительном (выборочном) просмотре документов фонда заслуженного для Беларуси, Литвы и Польши рода Радзивиллов нам стало ясно, что основной массив документов требует полистного просмотра дел и только после этого — составления аннотаций. Процент описания дел по заголовку в описи будет мал.

Как уже упоминалось в печати, примерно 30 тысяч дел Несвижского архива хранятся в Беларуси, остальные — в Польше, Литве, Украине. Как один из вариантов, возможно при их описании объединение ряда вопросов по группам, например: организационные вопросы; производство; отрасли сельского хозяйства; строительство и т. д., а далее их описание может строиться по группам — с помощью системы дескриптов. Немаловажно, чтобы последние легко поддавались распределению по тематическим группам и сохраняли свои связи, чтобы в итоге у нас не получился бессмысленный набор.

Состояние НСА к фондам Несвижского архива в настоящее время явно не может удовлетворить информационные потребности пользователей. Поэтому для обеспечения возможности полноценной работы исследователей с радзивилловскими документами, сокращения временных и финансовых расходов исследователей на поиск интересующей их информации необходимо переработать, а фактически создать заново НСА фонда. Очевидно, что целесообразно создавать электронный вариант НСА как наиболее соответствующий современным требованиям к оперативности и достоверности подготовки информации.

2. Некоторые способы и методы создания электронного НСА

Как уже говорилось ранее, наиболее полные возможности по поиску информации предоставляют системы, содержащие полные тексты документов. Технология реализации систем подобного типа представлена следующими магистральными направлениями:

- ✧ технология электронного архивирования, использующая возможности современных промышленных систем управления базами данных (СУБД);
- ✧ технология, основанная на специализированных системах индексирования и поиска.

Вторым направлением (или первым, если вести отсчет от начала использования в практической работе) является направление по созданию традиционных информационно-поисковых систем, основанных на т. н. атрибутивном поиске структурированных данных. Подобные системы автоматически повторяют структуру традиционных архивных справочников (описи, указатели, тематические каталоги и картотеки и т. п.). Определенные преимущества при создании таких БД очевидны. Так, значительно упрощается и ускоряется процесс поиска, пользователь может осуществлять многоаспектный поиск (сразу по нескольким атрибутам или реквизитам) и готовить отчеты различных видов.

Нам представляется наиболее целесообразным выбрать второе направление. Это позволит сократить финансовые расходы и время на реализацию проекта.

При построении традиционных информационно-поисковых систем особенно важное значение имеет определение системы дескрипторов-описателей. От качества разработки системы дескрипторов зависят как эффективность, так и сама возможность поиска информации. В ходе описания дел должен быть сформирован словарь дескрипторов — тезаурус, который представляет собой результат семантического и лексикостатистического анализа текстов документов. В тезаурусе перечислены термины (дескрипторы) и синонимичные им ключевые слова, показан метод устранения синонимии,

омонимии, полисемии, выражены родовидовые и ассоциативные связи между дескрипторами.

Основные принципы построения тезауруса:

- ✧ однозначность — должны быть устранены синонимия, омонимия, полисемия терминов;
- ✧ нормативность — включению подлежит только терминология;
- ✧ децентрализация — тезаурус строится «снизу вверх», от лексической единицы к тематической группе;
- ✧ фиксация объективно существующих отношений;
- ✧ независимость тезауруса от конкретной реализации системы;
- ✧ логическая чистота структуры тезауруса — недопустимость смешивания отношений разных видов;
- ✧ обратимость отношений (например, если есть родовое отношение, то должно быть и обратное ему видовое);
- ✧ открытость — допущение включения новых тематических классов;
- ✧ корректировка — возможность внесения новых терминов или их исключения;
- ✧ совместимость.

Разработка тезауруса осуществляется поэтапно:

- ✧ исследование лингвистических особенностей соответствующей отрасли знания;
- ✧ определение тематической области, которая должна найти отражение в тезаурусе;
- ✧ выявление пересекающихся тематических отраслей и разработка структуры тезауруса;
- ✧ изучение терминологии отрасли для каждой из тематических областей;
- ✧ разработка лексико-семантической основы тезауруса и распределение терминов в соответствии с его структурой;
- ✧ выявление синонимии, омонимии, родо-видовых и других отношений и фиксирование их в структуре тезауруса.

3. Требования к создаваемой информационной системе

Первое условие, которому должна отвечать создаваемая информационная система, — возможность удобного для пользователя-оператора ведения (ввода, изменения, дополнения, удаления) расширенного описания дел, которое включает в себя два типа характеристик: обязательные элементы и дескрипторы.

Заголовок должен содержать следующие обязательные характеристики:

- ✧ справочные данные (количество листов, крайние даты дела);
- ✧ поисковые данные (номера фонда, сдаточной описи, дела по сдаточной описи);

✧ данные о составе и содержании документов дела (аннотация, являющаяся аналогом полного заголовка дела).

Кроме обязательных элементов, используются также дескрипторы различных типов:

- ✧ тематические — перечень терминов, отражающих отдельную тему (предмет, вопрос, явление, событие, факт);
- ✧ географические — перечень географических названий;
- ✧ видовые — перечень видов документов;
- ✧ структурные — перечень названий организаций, структурных частей фондообразователя;
- ✧ именные — перечень имен.

Вторым требованием является наличие развитого поискового аппарата (языка), позволяющего исследователю достаточно просто описать то, что он хочет найти. Так как мы выбрали в качестве базового варианта классическую информационно-поисковую систему, то, как отмечалось выше, для нее характерен атрибутивный поиск информации, т. е. поиск информации по дескрипторам в структурированном массиве данных. Следовательно, задание значений поисковых полей (дескрипторов) для пользователя должно быть максимально упрощено (набор значений из списка, позиционирование на нужное значение по первым введенным буквам и т. д.).

Необходимо также, чтобы система имела надежные средства поддержания собственной работоспособности (копирование, восстановление после сбоев, надежность работы в сети и пр.), а также открытую для администраторов системы структуру данных, позволяющую им выполнять сервисные функции в системе и, при необходимости, получать данные для анализа работы системы (статистика обращений, количество введенных данных и др.).

На основании вышеизложенного становится очевидной необходимость проведения подготовительного этапа в реализации данного проекта, в ходе которого должны быть решены многие из обозначенных ранее проблем. Например:

✧ подготовка методики описания информации — составление макета тезауруса, состоящего из тематического, географического, именного, видового и, возможно, иных разделов, обуславливаемых составом вводимой информации и схемой систематизации фонда. Построение иерархической структуры (рубрикатора) документов фондов Радзивиллов;

✧ определение на основе избранной ранее информационной системы и системы программного обеспечения, порядка проведения работ по вводу информации и формированию электронного НСА, а также другие вопросы как организационного характера, так и информационно-методические.