

# ТЕХНОЛОГИЯ РАБОТЫ С ЭЛЕКТРОННЫМ СЕМЕЙНЫМ АРХИВОМ

И. М. Адамович, О. И. Волков, Н. А. Маркова

*Институт проблем информатики Российской академии наук*

Поступила в редакцию 26.02.2012 г.

**Аннотация.** Анализируются основные характеристики электронного архива семейной истории. Предлагается метод комплексной систематизации семейного архива, а также основанная на нем концептуальная модель хранения данных. Рассматриваются функции инструментария поддерживающей информационной технологии.

**Ключевые слова:** семейная история, документ, электронный архив, систематизация, объектная модель, эффективный доступ.

**Annotation.** The main characteristics of digital based family archive fund are considered. A method of comprehensive systematization is proposed. A conceptual model of metadata storage and supporting tools functionality are provided.

**Key words:** family history, digital archive, systematization, objects model, efficient access.

## 1. ВВЕДЕНИЕ

В процессе жизни и деятельности каждого человека образуется множество документов, причем большая часть из них имеет отношение и к другим лицам, прежде всего, к членам его семьи. Современные информационные технологии, как технически, так и экономически позволяют владельцу персонального компьютера сохранить электронные копии этих документов в индивидуальном хранилище. Однако для того, чтобы эффективно работать с накопленным материалом требуются специальные системные решения, как методические, так и инструментальные.

Феномен архивного фонда личного происхождения давно освоен профессиональными архивистами [1]. В последнее время все большее внимание даже в рамках государственных архивов уделяется архивам «рядовых» граждан [2]. Активно подключаются к сохранению личных архивов некоммерческие организации. Образованная в 1988 году независимая общественная организация «Народный Архив» [3], например, хранит материалы о жизни рядовых граждан. Имеется также множество примеров коммерческих предприятий, предлагающих спектр услуг по созданию и ведению семейных (персональных) электронных архивов.

Если даже «большая» история интересуется «источниками нового типа» – следами повсед-

невного опыта обычных людей, тем более важно их сохранение для истории «малой» – истории своего рода, истории своей семьи, личной истории.

Несложно оцифровать и сохранить старые документы, например, письма с фронта бабушки, мамин школьный дневник, себя в групповом снимке первоклассников. Еще проще сохранить поток цифровых фотографий сегодняшней жизни. Проблемы возникают с доступом к этой информации, поиском и представлением данных. Существующие методы и инструментальные средства решения этих проблем имеют ряд существенных ограничений, которые подробно были рассмотрены в [4].

В работе анализируются характеристики семейного архивного фонда, размещаемого в индивидуальном электронном хранилище, и предлагается концептуальная модель информационной технологии, на основе которой возможна комплексная систематизация документов, и которая может служить основой эффективного доступа к ним.

## 2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СЕМЕЙНОГО АРХИВНОГО ФОНДА

Требования к характеристикам и функциям инструментария, обеспечивающего формирование и ведение семейного архивного фонда (САФ) были подробно проанализированы нами в работе [4]. Анализ проводился на основе рассмотрения содержания архива и типовых задач,

с ним связанных. Перечислим основные характеристики САФ обобщенно.

1. Содержание САФ составляют: официальные биографические документы (свидетельства о рождении, браке, смерти, паспорта, удостоверения, справки, характеристики, трудовые книжки, дипломы, анкеты, автобиографии и т.п.), дневники и воспоминания, переписка, материалы служебной, общественной, творческой деятельности (планы, рукописи, черновики, проекты, публикации, повестки, извещения, протоколы и т.п.), имущественно-хозяйственные документы, изображения (фотографии, рисунки) и т.д.

2. Масштаб САФ (примерный порядок величин):

- Число изображений ~ 1 000–100 000
- Число изображаемых лиц ~ 10–1 000
- Число прочих объектов (животные, машины и т.п.) ~ 10–1 000
- Число изображаемых мест ~ 10–1 000
- Число факсимиле ~ 100–1 000 (~1–10 страниц на документ)
- Число разновидностей документов ~ 10–1 000

3. Локальность САФ. Хранимая в САФ информация содержит персональные данные ныне живущих людей, носит конфиденциальный характер и защищена Законом [5].

4. Связность изображений, относящихся к сложным факсимиле (учет надписей на оборотной стороне старых фото, многостраничных документов).

5. Вариантность представления. Из физически единственного экземпляра хранения документ должен попадать в любое число «виртуальных» альбомов, характеризующихся критерием отбора, порядком расположения (группировка, сортировка), способом представления.

6. САФ должен быть систематизирован на наиболее удобным для конкретного пользователя образом.

7. Открытость САФ. Как процесс накопления документов, так и процесс их систематизации протяженны во времени (практика «сдачи в архив», когда в традиционный архив в определенный момент времени сдается готовый набор документов, для САФ неприемлема).

Остановимся подробнее на вопросах систематизации САФ, как ключевых для создаваемой технологии.

### 3. СИСТЕМАТИЗАЦИЯ САФ

Нетрудно обозначить основные направления систематизации – это время, содержание документа и форма его представления. Значительно сложнее систематизировать конкретную совокупность документов.

Даже время не является однозначным параметром систематизации. Так «время» выпуска для документа может не совпадать с временным периодом (периодами), о которых идет речь. В частности, справка, выданная в некий момент, может относиться к некому периоду в прошлом. Таким образом, наряду со временем выпуска необходимо фиксировать референтные периоды времени, к которым относится содержание документа.

В традиционных архивах содержание документов (помимо основного деления фонд-опись-дело) систематизировалось с помощью указателей: хронологического, именного, предметного, географического. Важнейшее преимущество цифровой технологии – возможность соотносить элементы этих указателей не только с документами, но и между собой. Тем самым обеспечивается возможность «косвенного» поиска по связям.

Обратим внимание на одно из стандартизированных определений документа: «Документ – объект информационного взаимодействия в социальной среде, предназначенный для формального выражения социальных отношений между другими объектами этой среды» [6]. Таким образом, документ может быть охарактеризован с помощью связанных с ним объектов, которые, в свою очередь, связаны между собой. В качестве объектов САФ выступают, прежде всего, люди (персоналии), места, возможно, организации, а также любые другие сущности, к которым имеют отношение хранимые документы, и которые владелец архива выделяет, как значимые. Невозможно перечислить все возможные виды объектов, встречаемые в различных архивах, к тому же такое перечисление будет избыточным для конкретных архивов. Мы предлагаем пользователю самостоятельно задавать нужные ему виды объектов, а также виды связей, допустимых для объектов данных видов. Например, объекты-персоналии связывают различные варианты не только родственных, но и социальных связей.

Для характеристики процессов существования/изменения объектов во времени чрезвы-

чайно полезно ввести понятие событие, как ограниченный во времени процесс, имеющий отношение к конкретным объектам. Примерами событий являются: Свадьба, Путешествие, Юбилей. Полагаем, что событие может включать подчиненные события и быть представленным несколькими документами. Например, событие «Поездка по Золотому кольцу» включает события: «посещение Суздаля», «посещение Владимира», «посещение Ростова», а также, возможно, отдельные фотографии, запечатлевшие путешественников в дороге.

Не отвергая идею тегирования – сопоставления сущностям (объектам, событиям, документам) ключевых слов, считаем необходимым ограничить свободу их формулировок. Для этого помимо использования словаря тегов (открытого для пополнения), считаем целесообразным ввести иерархические конструкции тегов. Например:

«Путешествия»

    «Путешествия/автобусные экскурсии»

    «Путешествия/лодочные походы»

    «Путешествия/пешие походы»

        «Путешествия/пешие походы/  
одного дня»

        «Путешествия/пешие походы/  
многодневки»

Заметим, что для семейных архивов характерна значительная гетерогенность, а также всевозможные «невозможные» сочетания характеристик формы представления документов (например, рисунок на полях газетной вырезки, снимок выпускника, держащего только что полученный диплом на фоне университета и

т.п.). Следовательно, необходимо обеспечить возможность множественной классификации.

Итак, в качестве основы систематизации документов САФ мы предлагаем использовать сочетание объектной модели, относящихся к ним сущностей, и открытой иерархической классификации, как самих документов, так и компонентов модели. При этом, как документы, так и компоненты модели полагаются (по возможности) хронологически определенными.

#### 4. КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ САФ

Для того чтобы реализовать предложенную систематизацию, необходимо хранить метаданные, описывающие «картину мира» конкретного владельца-пользователя САФ. Построим такую модель обобщенно в терминах диаграммы Сущность-Связь (ER) – рисунок 1.

Под объектом САФ, будем понимать персоналию, семью, место, организацию или любую другую сущность, которую пользователь САФ соотносит с документами. Все объекты имеют имена, неформальные тексты – описания, а также сведения о времени существования. Для персоналии – это время жизни, для организации – период между основанием и ликвидацией. Временем существования характеризуются и населенные пункты.

Между объектами существуют (и могут быть определены) связи, различных видов. Полагаем, что в зависимости от потребностей пользователь сам задает их номенклатуру. Такие связи множественны: один и тот же человек может приходиться пользователю и однокашником и коллегой.

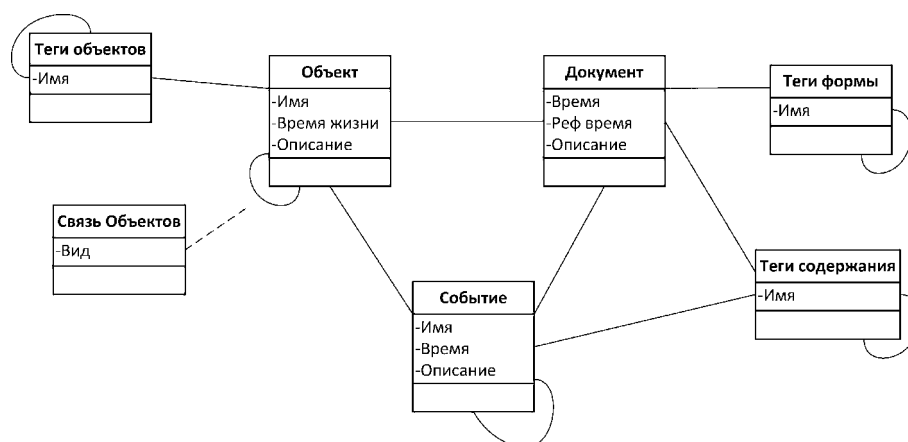


Рис. 1. Концептуальная модель САФ

Выделенная группа участвующих объектов в определенный промежуток времени может быть соотнесена с событием, имеющим свое описание. События могут образовывать иерархию.

Как объекты, так и события связываются с документами. Для документа определяется время выпуска и, возможно, референтное время – в общем случае, несколько периодов. Кроме того, документ оснащается неформальным описанием.

Наконец, объекты, события и документы классифицируются с помощью ключевых слов – тегов. Полагаем, что словарь тегов открыт, и теги в нем образуют иерархию. Мы разделяем три словаря тегов: объектный, содержательный и формы документов.

Теги содержания представляют некоторую пользовательскую классификацию, например, виды деятельности. Теги объектов – классификацию объектов, удобную для данного пользователя. Теги формы документа задают его знаковую систему (текстовой, изобразительный, идеографический), способ изготовления (фотографический, рукописный, печатный), распространение, и т.п.

Представленная концептуальная модель определяет содержание САФ и является ядром технологии работы с ним.

## 5. ФУНКЦИИ САФ

Основные функции САФ заключаются во вводе, хранении, организации доступа к документам и сопровождающей метаинформации, а

также поиске и различных вариантах представления хранимых данных. Полагаем, что загрузка документов проводится многократно, процесс систематизации, включающий составление/уточнение компонентов объектной модели, а также определение словарей тегов перманентен. Многократно в соответствии с сохраненными критериями отбора производится выборка документов – формирование альбомов.

Рассмотрим основные функции САФ (рис. 2), как операции над базой метаданных.

Ввод/редактирование данных об объектах, событиях, их связях, а также классификаторов – словарей тегов осуществляется с помощью диалоговых форм, представляющих необходимые поля (имена, даты, описания и т.д.).

Исходные файлы загрузки – это результаты сканирования, а также данные, переписываемые из памяти цифрового устройства (например, фотоаппарата), полученные по электронной почте или скаченные по сети. Мы полагаем, что в качестве документов используются не только изображения, но и произвольные мультимедийные объекты, в частности, аудио и видео материалы (поэтому термин «Изображение» в рисунке 2 весьма условен).

Загрузка файлов, отображаемых в режиме просмотра, сопровождается начальной систематизацией – вводом метаинформации, как касающейся текущего документа, так и его привязке к объектам и событиям, тегам. Файлы документов сохраняются в хранилище САФ с некоторыми условными именами.

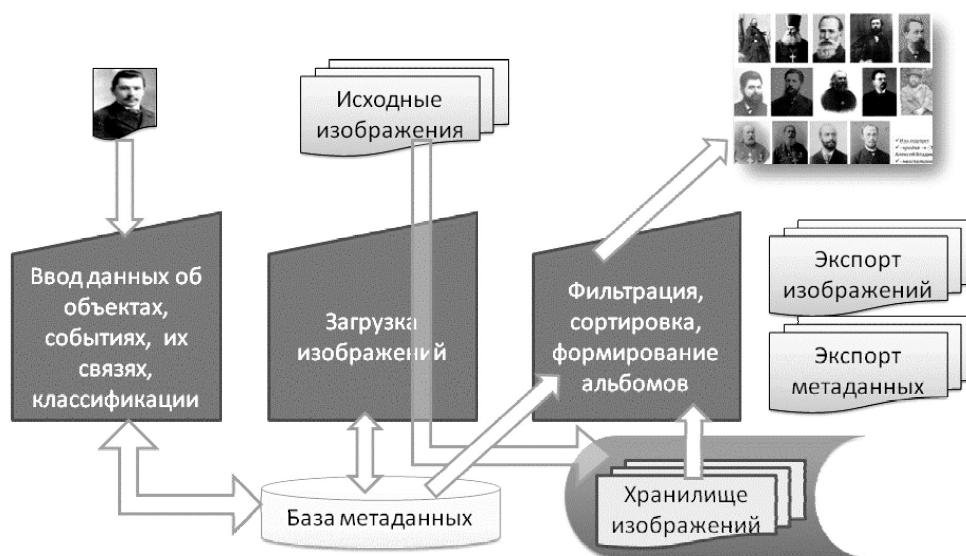


Рис. 2. Функции САФ

Лишь современные цифровые камеры обеспечивают автоматизированную фиксацию времени снимка. Для большинства старых фотографий, тем более сканируемых документов время необходимо задавать вручную. Тем более ручного ввода требуют референтные периоды.

Основа доступа в САФ – навигация и поиск материалов реализуется в терминах концептуальной модели, обеспечивая при этом возможность, как передвигаться по связям, так и задавать атрибуты сущностей в поисковой форме соответственно.

Благодаря многоаспектной классификации, включающей как теги, так и связи с объектами, можно формулировать сложные поисковые запросы, опираясь на легко формулируемые условия. Например, искать «официальные» (тег) документы, связанные с «родней» (вид связи) данного лица, в данный период. Вместе с этим возможен полнотекстовый поиск в именах и в описаниях.

Поисковый запрос вместе с параметрами представления, включая группировку и сортировку, в том числе ручную, сохраняются в именованные совокупности параметров, определяющих альбомы (аналоги виртуальных папок). Примерами альбомов являются «Походы в студенческие годы» или «Друзья отца». Заметим, что сохранение критерия отбора, вместо ссылок на конкретные отобранные документы, означает, что при новом обращении в альбоме, возможно, окажутся и другие документы, которые мы догрузили в архив или классифицировали соответствующим образом, что обеспечивает его актуальность.

Некоторые критерии отбора документов в альбомы являются очевидными и часто используемыми. Такие критерии (и, соответственно, связанные с ними альбомы) предопределяются заранее. Примерами являются документы, связанные с

- данным объектом (в частности, персоналий),
- данным событием,
- данным тегом (ключевым словом)

Удобным интерфейсным приемом является вынесение в общие сведения о данной сущности числа связанных с ней документов, щелчок по которому и может служить сигналом к открытию (визуализации) соответствующего предопределенного альбома.

Более сложные критерии отбора формулируются как логические сочетания простых критериев,

а также условий, обуславливаемых содержанием объектной модели. Например, «найти документы, связанные с объектами, связанными данными отношениями с данным объектом».

Альбом визуализируется в двух вариантах. В первом «значки» (preview) документов или стопки документов, представляющих событие, или сложный документ, составляют таблицу. В зависимости от потребности пользователя стопки раскрываются или остаются сложенными. Во втором варианте – отображается по одному полноценному изображению с возможностью масштабирования и перехода к предыдущему/следующему изображению.

В каждый текущий момент в открытом альбоме одно из изображений является текущим, выделенным. Для выделенного изображения отображаются свойства, связанного с ним документа, в том числе, его связи с объектами. Переходя по этим связям, пользователь, возможно, открывает новые альбомы или просматривает цепочки отношений между объектами.

Альбомы являются средством не только просмотра САФ, но и отбора изображений для экспорта в другие приложения. В рамках САФ не имеет смысла реализовывать развитый аппарат работы с изображениями: подготовки презентаций, художественного оформления, восстановления утрат (ретуши) – для всего этого можно найти соответствующие готовые приложения. САФ же обеспечивает удобный механизм поиска нужных изображений, которые по специальной команде в виде файлов с указанными именами копируются в указанную папку.

Экспорт и импорт метаданных целесообразно обеспечить в вариантах, как дополнительные модули, набор которых определяется потребностями пользователя. Один из наиболее востребованных вариантов экспорта нужен для передачи данных в генеалогические программы, реализующие, в частности, построение родословных древ. Для его реализации предполагается использовать наиболее распространенный формат обмена генеалогическими данными GEDCOM [7]. Импорт метаданных позволит использовать внешние данные, возможно, уже подготовленные стандартными инструментами.

## 6. ЗАКЛЮЧЕНИЕ

В работе предложена модель информационной технологии САФ, позволяющей систематизировано хранить и использовать большой

объем личных документов. В основу систематизации положена объектная модель, связывающая документы с персоналиями и другими сущностями, а также изображаемые (упоминаемые) сущности друг с другом.

Предложенная модель обладает рядом особенностей, отличающих ее от моделей, на которых основаны современные программные средства. Она сочетает единое пространство хранения документов, обеспечивающее их многократное использование в различных контекстах, характерное для стандартных средств современных операционных систем, с опорой на персоналии и межличностные связи, используемых для классификации документов, свойственной генеалогическим программам. Причем помимо родственных связей в рассмотрение включены и любые другие, в частности, социальные связи, а помимо персоналий любые объекты.

В настоящее время разрабатывается прототип инструментальных средств, реализующих предложенную технологию. В рамках пробной эксплуатации прототипа будут отработаны интерфейсные решения, обеспечивающие для пользователя удобный доступ к функциональным (а, в конце концов, и к концептуальным) возможностям.

Предложенная технология персонально ориентирована: как объектная модель, так и классификаторы в ней создаются пользователем. С одной стороны, это предоставляет самый широкий спектр возможностей, с другой, сопряжено с методическими трудностями (особенно для людей далеких от информационных технологий). По мере накопления опыта использования инструментария будут вырабатываться

методические решения и рекомендации, которые должны стать важным компонентом технологии САФ.

#### СПИСОК ЛИТЕРАТУРЫ

1. Правила организации хранения, комплектования, учета и использования документов архивного фонда Российской Федерации и других архивных документов в государственных и муниципальных архивах, музеях и библиотеках, организациях Российской академии наук. Утверждены 18 января 2007 г. №19. – URL: <http://www.rusarchives.ru/lows/pohkuidaf.shtml#51>

2. *Шпагина М. П.* Личные фонды: трансформация феномена // Документ в системе социальных коммуникаций: Сборник материалов III Всероссийской научно-практической конференции с международным участием. – Томск: Томский государственный университет, 2008. – С. 152–155.

3. *Добрушкина Н. И., Илизаров Б. С.* Новая идеология информационно-поисковой системы по документам архивов // Архивы образа и звука. М.: Центр документации “Народный архив”, 1996. С. 41-53; URL: <http://old.nffedorov.ru/na-nca/na/61i.html#a4>

4. *Адамович И. М., Волков О. И., Маркова Н. А.* Организация семейного архивного фонда // Труды 13-й Всероссийской научной конференции “Электронные библиотеки: перспективные методы и технологии, электронные коллекции” (RCDL’2011). – С. 139–145.

5. Федеральный закон Российской Федерации от 27 июля 2006 г. № 152-ФЗ “О персональных данных” // Российская газета. Федеральный выпуск № 4131 от 29 июля 2006.

6. ГОСТ Р 52292-2004 – «Информационная технология. Электронный обмен информацией. Термины и определения». – М.: 2005.

7. GEDCOM [Электронный ресурс]: Материал из Википедии – свободной энциклопедии – URL: <http://ru.wikipedia.org/wiki/GEDCOM>.

**Адамович Игорь Михайлович** – зав. отделом ИПИ РАН, к.т.н., e-mail: [adam@amsd.com](mailto:adam@amsd.com)

**Волков Олег Игоревич** – ведущий программист ИПИ РАН, e-mail: [volkov@amsd.com](mailto:volkov@amsd.com)

**Маркова Наталья Александровна** – вед. научн. сотр. ИПИ РАН, к.ф.-м.н., e-mail: [nMarkova@ipiran.ru](mailto:nMarkova@ipiran.ru), тел.: +7 (916-697-2250).

**Adamovich Igor** – department head, Institute of Informatics Problems, Russian Academy of Sciences (IPI RAN), candidate of Science (PhD) in engineering, e-mail: [adam@amsd.com](mailto:adam@amsd.com)

**Volkov Oleg** – senior software developer IPI RAN, e-mail: [volkov@amsd.com](mailto:volkov@amsd.com)

**Markova Natalia** – candidate of Science (PhD) in physics and mathematics; senior researcher IPI RAN e-mail: [nMarkova@ipiran.ru](mailto:nMarkova@ipiran.ru), tel.: +7 (916-697-2250).