

Электронные биографические ресурсы

© Н.А. Маркова, И.М. Адамович

Институт проблем информатики Российской академии наук, г. Москва
nmarkova@ipiran.ru

Аннотация

Рассматриваются электронные коллекции, содержащие биографическую информацию. Дается обзор их современного состояния, проблем и перспектив. Анализируются факторы, влияющие на эффективность биографического исследования, использующего электронные ресурсы.

1 Введение

Изучением биографий традиционно занимались профессионалы-историки, в частности, историки науки. Биографии группы лиц, объединенных родством, изучали генеалоги, а лиц, связанных некими социальными связями, – просопографы. По мере роста общественного интереса к генеалогии и «складывающегося в обществе, а вместе с ним и в науке, приоритета личности» [1], к этой работе стали подключаться архивные службы и библиографы, выполняющие заказы пользователей архивов и библиотек по поиску биографической информации. Новый потенциал средств хранения и распространения вместе с новыми коллективными формами формирования электронных биографических ресурсов способствует тому, что такую работу вскоре сможет выполнить фактически любой человек, знакомый с азами информационных технологий.

Количество электронных ресурсов, хранящих биографическую информацию: энциклопедий, словарей, генеалогических и просопографических баз данных, а также «персоналий» в коллекциях произвольного назначения – растет стремительными темпами. Возможность сохранения больших объемов информации вместе с возможностью организации удаленного доступа к ней для широчайшего круга пользователей сыграли стимулирующую роль для удовлетворения объективно существующей потребности в изучении и распространении сведений биографического характера.

Однако большая часть биографической информации до сих пор пылится в архивах. В начале 1990-х был открыт доступ к материалам отечественных архивных фондов, ранее закрытый для исследователей. Пользование архивами, однако, представляет

непростую задачу. Как правило, их каталоги и описи на бумажных носителях плохо систематизированы, а подчас и плохо читаемы за ветхостью. Заказанные материалы попадают в руки читателя не сразу. Иногда заказ нужно ждать целую неделю. Материалы могут быть весьма объемны, что при отсутствии индексации приводит к непростой длительной работе по их просмотру в поисках нужной фамилии.

Следующий большой шаг, декларируемый «Концепцией информатизации архивного дела России» [6], – обеспечение права граждан на информацию, предоставление им удаленного доступа к информационным ресурсам архивов. А для этого информация традиционных архивов должна быть оцифрована. Заметим, что в США национальный архив (<http://www.archives.gov>), в котором любой желающий может найти интересующую его биографическую информацию, давно оцифрован.

Архивные данные представляют только сырой материал для биографии, даже если речь идет не о развернутом литературном жизнеописании, а о простой биографической справке – «персоналии». Для ее составления требуется собрать и обобщить биографические факты, под которым понимается «высказывание..., являющееся ответом на вопросы типа кто?, что?, когда?» [3, с. 53], упорядочить их определенным образом, связать между собой и с внешними объектами. Объединение таких персоналий по определенному принципу может служить основой генеалогического или просопографического анализа или же быть базой информационно-справочного аппарата.

К сожалению, существующие хранилища биографических данных, будь то архивные данные, электронные публикации печатных изданий или же создаваемые сразу в электронном виде коллекции персоналий не соответствуют гигантскому потенциалу современных информационных технологий. Большинство из них – это текстовые («плоские») или слабо размеченные документы. Но и в более развитых инструментальных средах – базах данных и «сильно» размеченных документах (например, в технологии wiki) – возможности организации доступа далеки от требований эффективного использования в научных, просветительских, социальных целях. В преддверии массовой оцифровки архивов и печатных источников особенно важно осознать, какими свойствами должны обладать биографические ресурсы, чтобы пользование ими было бы эффективным.

Труды 12^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

Рассмотрим основные виды существующих ныне биографических электронных ресурсов, начиная с их бумажных предшественников. Перечислим типовые задачи, решаемые с их помощью, и их основные характеристики. Будем иметь в виду, прежде всего, потребности представления ретроспективных биографий – людей из прошлого.

2 Источники биографических информационных ресурсов

Начнем наше рассмотрение с традиционных бумажных ресурсов. Информационные ресурсы, содержащие биографическую информацию, многообразны и различаются по нескольким направлениям.

Историки-профессионалы выделяют несколько вспомогательных исторических дисциплин, имеющих дело с биографическими данными: биографика (изучение индивидуальных биографий), генеалогия (реконструкция истории рода в биографиях его членов) и просопография (изучение коллективных биографий социальных групп). С точки зрения справочно-информационного использования «уже готовых» биографических ресурсов различия между науками, применяемыми при их создании, не так существенны: для получения генеалогической информации можно воспользоваться энциклопедией, а биографическую справку можно найти в родословной росписи.

Биографические данные содержатся в специально написанных текстах-биографиях (индивидуальных, коллективных или семейных) и в формальных документах учета, а также их можно извлечь из произвольных информационных объектов, посвященных какому-либо объекту, явлению, процессу, к которым имело отношение изучаемое лицо (лица).

Биографии существенно различаются по объему: от пары строк до нескольких томов. Начиная с 19 века, широко публикуются словари, справочники, содержащие биографические справки как части универсальных энциклопедий или как независимые издания. На сайте Российской государственной библиотеки можно найти указатель «Русские биографические и биобиблиографические издания» (http://www.rasl.ru/e_resours/SBO_Guide-Golobko/start.html), содержащий сведения о более чем 400 изданиях такого рода.

Если в энциклопедии попадают сведения о лицах, в той или иной степени «выдающихся», то сведения обо всех прочих, равно как и дополнительные сведения о вошедших в энциклопедии лицах, можно почерпнуть из архивных и опубликованных в печати документов учетного характера. Архивные «учетные» документы – это метрические книги, исповедные ведомости, ревизские сказки, дворянские родословные книги и т. п. В справочнике [4] перечислено несколько сотен видов архивных документов, содержащих персональные данные.

Опубликованные (в печати) персональные данные – это книги памяти, адрес-календари, списки учащихся и т. д. К ним относятся как непосредственно

издаваемые сведения, так и публикация архивных документов [10]. Большая работа по публикации источников (в частности, «переписных книг» – аналогов современной переписи, «служебных книг» с записями о назначениях на должность) проводилась в последней трети девятнадцатого века. В отечественной практике интерес к таким публикациям возродился лишь в самом конце двадцатого века.

Традиционно одной из наиболее заинтересованных в биографической информации категорий лиц являются библиографы. Бумажная «Картотека персоналий – библиографическая картотека, отражающая документы, связанные с жизнью и деятельностью отдельных лиц, и организованная в алфавитном порядке их имен» составляет неременный компонент системы каталогов библиотеки. Сведения персонального характера составляют часть библиографической информации, фиксируемой «Российским коммуникативным форматом» [9]. В соответствии с ним Предметная рубрика – Персоналия должна содержать краткую биографическую справку (именную группу, идентифицирующий признак, например, писатель, даты жизни). Таким образом, каталоги библиотек и архивов представляют важнейший биографический ресурс, а их сотрудники выполняют функции биографов. С другой стороны, при подготовке биографий пишущих лиц (ученых, литераторов) важнейшим компонентом является биобиблиографический список трудов – еще одна забота библиографа.

Чрезвычайно важным и чрезвычайно сложным является вопрос о соотношении различных ресурсов, представляющих данные одной биографии. За исключением первичных документов, остальные представляют собой результат цепочки аналитико-синтетической переработок. Согласно работе [1], подготовка биографической справки включает процессы осмысления, структуризации, свертывания, формализации. При этом выявление фактов осуществляется на базе первичных источников, не подвергнутых авторской интерпретации или содержащих минимум ее. Ссылки на первоисточник получения информации, а также список использованной литературы обязательны для обеспечения проверяемости справки.

Совсем другие методы используются в традиционной биографике, конечным результатом которой является «составление повествования» [3]. Оно включает философское, психологическое (возможно, и художественное) осмысление жизненного опыта; дает картину жизни человека, развитие его личности в связи с общественными обстоятельствами эпохи. Предполагается, что биограф интерпретирует факты и заполняет информационные лакуны в архивных документах, исходя из собственной интуиции или внешних по отношению к изучаемой личности явлений.

Наконец, третий вариант составления биографии, к сожалению, все более распространяемый по мере развития средств коммуникации (сопряженно-

го с отсутствием редактуры), – это домысливание фактов, не опирающееся ни на надежные источники, ни на серьезное осмысление.

3 На пути к электронным биографическим ресурсам

Уже при переходе от архивных материалов к их книжной публикации существенно возрастает доступность информации. Магистральный путь обеспечения эффективного доступа к информационному ресурсу (а также его безопасного хранения) – перевод его в электронную форму. Бумажные библиотеки и архивы хранят колоссальное количество биографических данных, перевод которых в электронный вид, как можно надеяться, – вопрос времени.

Формирование электронных биографических ресурсов идет как путем публикации существующих бумажных материалов, так и путем создания новых исходно цифровых коллекций, элементы которых складываются из более чем одного бумажного (или электронного) источника.

В табл. 1 перечислены характерные примеры электронных публикаций архивных материалов и книжных изданий.

В простейшем случае электронная публикация сохраняет визуальную копию документа-источника или/и результат его распознавания (автоматического или ручного). Большинство из перечисленных примеров представляют текст – результат распознавания (с небольшим добавлением гиперссылок). Исключениями являются ресурсы, сохраняющие визуальную копию в виде графических файлов (постранично сканированные книги): 1.b (формат DeJaVu) и 3.a (формат jpg), а также ресурсы 2.a, 4.b – результаты более глубокой переработки – это базы данных. Новые биографические ресурсы, как правило, в большей степени используют преимущества машинной информационной технологии, информация в них представляется либо в виде баз данных, либо с использованием развитого аппарата разметки текстов – в данные включены элементы, которые позволяют их интерпретировать.

В формальных биографиях людей определенной социально-культурной страты может быть наведен некоторый общий порядок: выявлены рубрики, подрубрики, элементарные характеристики, принимающие значения из определенной области. Такого рода характеристики допускают машинную интерпретацию, к ним относятся: время, место событий, связи с другими лицами или организациями. В традиционном делопроизводстве формальные биографические данные – это заполненные анкеты.

В общем случае в биографиях присутствуют компоненты, принципиально неформализуемые: философское, психологическое осмысление жизненного опыта.

Наконец, определенная часть биографической информации может быть представлена в терминах специальной, разработанной для данной страты лиц классификации. Разработка таких классификаций – сложная задача.

Биографические базы данных основаны на хорошо формализуемых частях биографий. Что не исключает присутствие в них неформализуемых данных в виде полей специального вида – текстов и изображений. На сегодняшний день широкое распространение получили два варианта «анкетных» баз данных: просопографические и генеалогические.

На протяжении трех десятилетий (1970 – 2000-е гг.) в отечественной исторической науке возник и бурно развивался жанр просопографических баз данных, создаваемых на основе массовых источников биографической информации. В работе [12] рассмотрено более сотни баз, содержащих демографические, социальные, образовательные и профессиональные характеристики групп лиц. Статья представляет биографические ресурсы, являющиеся инструментами профессиональных исследователей. Однако за последние годы в области просопографических баз данных произошел качественный сдвиг. Касается он как масштаба хранилищ – сегодня он измеряется миллионами лиц, так и категориями потребителей – большинство из которых – непрофессионалы.

Современные российские и зарубежные генеалогические базы данных рассмотрены в работе [8]. Здесь также речь идет о миллионах лиц, а использование – часто не только для получения справок, но и для пополнения – предполагается любым желающим.

Биографические ресурсы, основанные на гипертекстовых документах, не требуют столь жесткой формализации данных, как базы данных. В предельном случае их элементы – это простой («плоский») текст. Однако в таких документах сложно искать информацию, в них отсутствует основа для машинной интерпретации данных.

В последние годы заметную конкуренцию реляционным базам данных в области биографических ресурсов (и не только в них) составили размеченные гипертекстовые технологии, в которых сочетается свобода работы со слабо формализуемыми данными с возможностью автоматизированной обработки их формальных частей. Наиболее бурное развитие переживает технология wiki. Основное преимущество ее заключается в том, что она позволяет реализовать коллективную работу с документами. Для биографических данных существенно значение имеет предоставляемая wiki свобода в совместном использовании формализуемого и неформализуемого знания. При этом сама формализация (в отличие от реляционных баз данных) может осуществляться по ходу накопления данных. Собираемый «с миру по нитке» материал легко редактировать, при этом существует контроль версий (всегда есть возможность вернуться к предыдущему варианту). Многократные итерации улучшают качество статей.

Таблица 1. Примеры электронных публикаций

Вид	Название	Адрес
1. Биографические словари	a. Большая Советская Энциклопедия (содержит тысячи биографических статей)	http://slovari.yandex.ru/dict/bse
	b. Русский Биографический Словарь Половцева	http://www.rulex.ru/xPol/index.htm
	c. Биографический корпус словаря Брокгауза и Эфрона	http://www.rulex.ru
2. Архивные документы	a. Боярские списки XVIII века	http://zaharov.csu.ru/bspisok.pl
	b. Перепись населения 1710	http://census1710.narod.ru/admin.htm
	c. Исповедные ведомости Пречистенского сорока	http://www.mtu-net.ru/sherbach/genialog/prech.html
3. Книжные издания массовых биографических данных	a. Вся Москва 1901	http://moskva.rodstvo.ru/
	b. Списки выпускников духовных учебных заведений	http://www.petergen.com/bovkalo/duhov.html
	c. Петров П. Н. История российского дворянства	http://russianfamily.ru/
4. Каталоги библиотек и архивов	a. Личные архивные фонды в государственных хранилищах СССР	http://rusarchives.ru/guide/lf_ussr/abc.shtml
	b. Персоны в Российском государственном архиве литературы и искусства	http://www.rgali.ru/search.do?objectType=60776607
5. Литературные биографии	a. Жизнь Замечательных Людей - 133 электронных книги (из 1000)	http://zsl.lib.ru/
	b. Вл. Соловьев. Судьба Пушкина	http://www.magister.msk.ru/library/pushkin/about/solov001.htm
	c. Выпускники МГУ в Новосибирском научном центре СО РАН. 1957-2007. - Новосибирск: Гео, 2007	http://www.prometheus.nsc.ru/elibrary/2007mgu/index.ssi
6. Неструктурируемые документы, содержащие биографические данные	d. Мемуары. Записки. Дневники в некоммерческой электронной библиотеке "ImWerden"	http://imwerden.de/cat/modules.php?name=books&pa=last_update&cid=35
	e. Документы личного характера	http://uni-persona.srcc.msu.su/site/ind_cont.htm

Но наряду с «продвинутыми» технологиями для биографических ресурсов, особенно любительских, имеющих тенденцию к неограниченному росту, все еще употребляются простейшие инструментальные средства. В результате ресурс представляется набором текстов.

Промежуточное место занимают такие хранилища биографической информации, как система «Просис» [12], где в рамках единого ресурса существует гипертекстовая (неформализуемая) часть и реляционная база данных, в которой хранятся формализуемые, однородные данные – «Анкеты».

Также как «хорошая» бумажная книга оснащается справочно-информационным аппаратом – именными, географическими и предметными индексами, списками источников и пр., «хороший» электронный биографический ресурс оснащается соответствующими механизмами, часть из которых – аналог соответствующих бумажных индексов, другая реализует специфически электронные возможности, определяемые функциями соответствующего инструментария. Таким образом, под электронным биографическим ресурсом мы понимаем совокупность его информационного содержания – основного и вспомогательного (справочного, систематизирующего), а также инструментария. Выделим среди электронных биографических ресурсов специаль-

ный класс – коллекции персоналий, наиболее интересный с точки зрения как применения информационных технологий, так и обслуживания основного потока биографического поиска.

4 Коллекции Персоналий

В практике электронных коллекций, содержащих биографические данные, широкое распространение получил термин «персоналия». Как указывалось выше, термин этот уже применялся в библиотечном и архивном деле. Используется он и в просографии [12]. Именно так называются содержащие биографические справки разделы порталов и сайтов учреждений, обществ, а также ресурсов, посвященных краеведению или историческим событиям.

К сожалению, четкого определения этого понятия пока нет, что подтверждает статья [11], в которой были проанализированы все имеющиеся словарные определения этого понятия и сделан вывод, что ни одно из них не соответствует практике его применения.

На практике «персоналиями» в электронных хранилищах называют биографические данные об индивидуумах (лицах, персонах), собранные в некоторую коллекцию («картотеку», список). Такого рода коллекции либо представляют раздел некото-

рого ресурса (портала, сайта, базы данных), либо существуют независимо. При этом отдельная «персоналия» либо представляет биографическую информацию, касающуюся персоны в целом, либо касается частных аспектов ее биографии. В первом случае – назовем его «общим» – персоналией будет являться и краткое «Иванов Г.И. (1876–1930)», и развернутая биография. Во втором случае – «частном» – определенный «биографический контекст» фиксируется на уровне коллекции персоналий, например, [Вся Москва 1901] – «Иванов Григ. Ив., куп. Газетный, д. Грацева, Металлич. изд».

Назовем «коллекцией персоналий» совокупность биографических сведений о группе лиц, отобранных в соответствии с определенными критериями. Нас интересуют конкретно электронные коллекции, т. е. информационные системы, позволяющие сохранять данные в цифровом виде и обеспечивающие доступ к ним через сети передачи данных. В частности, необходим доступ к биографическим сведениям, касающимся конкретного лица, которые мы и будем называть персоналией.

Вводя термин «коллекция персоналий», мы избавляемся от привязки к исходным назначениям биографических информационных ресурсов на пути к обсуждению вариантов их возможного использования. В действительности коллекция персоналий, изначально представлявшая «генеалогическую» направленность, используется для получения биографической справки, а просопографическая база или энциклопедический словарь предоставляют материал для составления семейного древа.

Коллекции персоналий различаются в следующих аспектах.

1. Отношение к источникам: электронные издания бумажных биографических источников или ресурсы, формируемые на основе множества бумажных источников или электронных ресурсов, а также, возможно, других видов информации.

2. Критерии отбора лиц: нет ограничений, «исторические личности», пространственно-временные и социально-политические ограничения; ограничения-связи с конкретными учреждениями, обществами, лицами, произведениями, событиями и т. п.

3. Ответственность за формирование: конкретные составители (лица и организации), сообщество пользователей или их различными сочетаниями (например, заявку дает пользователь, а редактирование и ввод данных осуществляют ответственные лица).

4. Биографический жанр: конкретные выделенные аспекты биографии, биографическая справка, хронология событий, жизнеописание – повествовательная биография и пр.

5. Возможности доступа, в том числе, навигация, поиск, визуальное представление, экспорт и импорт.

В табл. 2 приведены типовые примеры коллекций, составленных более чем из одного источника. Критерий отбора лиц вытекает из названия (или комментария). Коллекции 1, 3, 7, 8 формируются совместным творчеством всех пользователей, ос-

тальные – результат деятельности авторского коллектива.

Возможности организации доступа к отдельным персоналиям или их группам вместе с возможностями их наглядного представления существенным образом зависят от уровня их систематизации. Основными аспектами систематизации являются структуризация, классификация и связывание данных.

Выделение коллекции персоналий из электронных биографических ресурсов основывается на предположении об определенной систематизации ресурса: речь идет о более чем одном лице, существует аппарат выборки, который вычлняет сведения о конкретном лице. Таким образом, к коллекциям персоналий не относятся неструктурируемые документы, содержащие биографическую информацию, а также индивидуальные биографии в случае их независимого хранения.

Если ресурсы, посвященные индивидуальным биографиям, содержат систематизированные сведения о группе лиц, связанных с основным лицом, то эту часть ресурса можно рассматривать как коллекцию персоналий. В данном случае основное лицо играет роль критерия отбора, схожую с участием в событии или членством в обществе. Реконструкция социальных связей – родственных, дружеских, соседских – одна из важнейших составляющих работы биографа.

Аналогично, неструктурируемые документы могут представлять компонент биографии лица из коллекции или быть особым, дополнительным объектом, на который ссылаются тексты персоналий.

Среди биографических ресурсов встречаются и такие, в которых персональные сведения сгруппированы не по лицам, а по общим для выделенного круга лиц признакам, например, персоналии физического факультета МГУ – <http://museum.phys.msu.ru/rus/books/pdf/physics-faculty.pdf>. В ресурсе представлены списки по должностям, званиям, наградам, премиям и пр. Если бы эти сведения представляли собой не электронную публикацию бумажной книги, а использовали возможности современных информационных технологий, то наряду с выборкой по признаку вполне возможно было бы организовать выборку по лицу. Наличие выборки по лицу будем рассматривать отличительной чертой коллекции персоналий.

В зависимости от степени универсальности охватываемого коллекцией персоналий информационного ресурса в качестве «дополнительных объектов» могут выступать различные сущности. Так, в биографиях художников – это картины, математиков – математические объекты. Фактически неограничена номенклатура дополнительных объектов в универсальных энциклопедиях (например, Википедии). Типичными дополнительными объектами являются учреждения (в том числе, образовательные), общества, исторические события, географические объекты.

Связывание – явное определение связей между персоналиями или между персоналией и объектом другой природы – важнейшее преимущество, даваемое электронным инструментарием перед бумажным. К сожалению, конкретные реализации персональных коллекций не всегда и не в полном объеме им пользуются.

Другой аспект систематизации персоналий – классификация, т. е. выделение некоторых (возможно структурно организованных) классов, категорий, к которым может быть отнесены данная личность или определенный аспект ее биографии. Соотнесенность с другим объектом или другой персоналией тоже может рассматриваться как специфическая форма классификации.

Таблица 2. Примеры коллекций персоналий

Название, характеристика	Масштаб	Адрес
1. Раздел «Персоналии» в Википедии (выдающиеся личности)	>100,000	http://ru.wikipedia.org/wiki/Категория:Персоналии по алфавиту
2. Хронос – «сведения обо всех исторических личностях за все века человеческой истории»	~30,000	http://hrono.info/biograf/index.php
3. Математическая генеалогия	>14,000	http://genealogy.math.ndsu.nodak.edu
4. Биографии отечественных художников	>42,600	http://artru.info/ar/
5. Барды	5,000	http://www.bards.ru/persons.php
6. Персоналии экономики, социологии и менеджмента	1,700	http://www.ecsocman.edu.ru/db/sectx/1111.html
7. Free Family History – международная генеалогическая база	600,000,000	http://www.familysearch.org
8. Всероссийское генеалогическое древо	>150,000	http://baza.vgd.ru/
9. Жертвы политического террора в СССР	>2,600,000	http://lists.memo.ru/
10. Защитники Отечества, погибшие и пропавшие без вести в Великую Отечественную войну (ОБД «Мемориал»)	>8,600,000	http://www.obd-memorial.ru/Memorial/Memorial.html

5 Типовые сценарии использования биографических ресурсов

Для того чтобы оценить содержание коллекции персоналий, а также предоставляемые ее возможности доступа, необходимо рассмотреть, какие типовые задачи решают потребители биографических ресурсов. В силу различий этих задач и критерии оценки будут различны.

Основные задачи пользователей коллекций персоналий, без ограничения общности, сформулируем следующим образом. Требуется получить (найти и извлечь) биографическую информацию:

- с познавательной, образовательной, справочно-информационной целью;
- с целью пополнения (усовершенствования) биографических ресурсов;
- как исходные данные биографического (просопографического, генеалогического) исследования.

Первый вариант охватывает как случаи простого любопытства, так и необходимость биографических справок в образовательном, исследовательском, художественном процессах.

Пополняемые биографические ресурсы (второй вариант) – это биографические словари и справочники (как электронные, так и бумажные) или разделы «персоналия» ресурсов различного назначения. В случае библиографических и археографических ресурсов задача формирования персоналий совмещена с задачей «предметизации» - процесса формирования предметных рубрик по терминологии библиографов.

В третьем случае пользователем является исследователь-профессионал или исследователь любитель. Это и ученые тех или иных исторических дисциплин, и краеведы-любители, и получающие все большее распространение службы, выполняющие генеалогические исследования по заказам клиентов. Проблемы справочно-информационного обслуживания, целью которого является реконструкция биографии, выполняемого силами библиографов, рассматривают работы [1] и [5]. Задачи, решаемые такого рода пользователями с привлечением коллекций персоналий, многообразны. В том их числе: выявление круга лиц, заполнение лакуны в их биографиях, сведение воедино данных из различных источников. В последнем случае, как отмечает [12, с. 123], в идеале требуется сведение всех собираемых данных в единый ресурс – «виртуальный метаисточник» при сохранении всех «исходных» данных для их последующего вероятного использования.

В первом случае, когда речь идет об известном персонаже, вполне вероятно хватит одного обращения к соответствующей (универсальной или специализированной) коллекции с запросом, если полученный ответ удовлетворит пользователя. В остальных случаях пользователь обращается к коллекции (коллекциям) многократно, на каждом шагу формулируя определенный запрос и получая ответ. При этом ищутся сведения об одном лице или о группе лиц, а параметрами поиска служат как точные идентификационные данные, так и отрывочные сведения. Типичны запросы по связям (родства, профессии, образования, проживания и пр.) с дру-

гими лицами или какими-то другими сущностями (например, организациями). Проблемы пользователя усугубляются тем, что даже если ему известны точные идентификационные данные и ориентировочное время жизни, в силу одноименности лиц, ответ будет неоднозначен.

Ответ – извлекаемая биографическая информация – характеризуется вариантами детальности содержания, включения в него тех или иных аспектов. Важнейшим результатом поиска служат отсылки к источникам (как электронным, так и традиционным – публикациям, архивам). Дальнейшее использование данных – визуальный анализ, автоматизированная обработка, поиск информации в указанных источниках. В первом случае существенное значение имеет наглядность визуального представления и инструментальные возможности уточнить полученные сведения или продолжить поиск, например, воспользовавшись гиперссылками. Во втором случае коллекция должна предоставлять возможность экспорта биографических данных в машиночитаемое представление. Наконец обращение к источникам может потребовать перемещения в библиотеку или архив. Результаты анализа данных очередного шага – одна из основ для формулировки нового запроса.

Суммарно неформально пожелания потребителя биографического ресурса сформулируем так: «за минимальное время найти и извлечь релевантную конкретной задаче [только нужную] информацию в максимально наглядной визуальной форме (или в конкретном формате экспорта)». Очевидно, что на успешность задачи влияют как средства формулировки очередного запроса и формы ответа, так и аппарат организации многошагового процесса поиска.

6 Характеристики биографических ресурсов

Даже беглый взгляд на существующие электронные биографические ресурсы демонстрирует весьма слабое использование в них преимуществ новых информационных технологий. Большая часть из них – это текстовые документы, с весьма незначительным использованием гиперссылок. Поиск конкретной биографической информации в них чрезвычайно затруднен.

С жестко структурированными, реализованными средствами реляционных баз данных коллекциями персоналий другая беда: дублирование данных одного и того же лица в нескольких персоналиях или объединение данных разных лиц в одной персоналии. Например, в ОБД «Мемориал» на фамилию Скундин находим 13 персоналий, при внимательном рассмотрении соответствующих всего четырем лицам.

Потребительские свойства коллекции персоналий зависят как от содержания, так и от инструментальных характеристик среды реализации. Под инструментальными характеристиками понимаются функциональные, интерфейсные возможности ин-

струментария. Однако для того чтобы воспользоваться инструментальным потенциалом, необходимо, чтобы содержание было бы хорошо систематизировано. Систематизация – информационно-аналитическая деятельность, в рамках которой производится частичная формализация знания, структуризация, классификация, связывание. Именно результаты систематизации являются базисом, на основе которого соответствующие инструментальные средства позволяют организовать эффективный доступ к информации.

Проиллюстрируем влияние всех трех составляющих на эффективность решения задач биографического поиска.

6.1 Содержательные характеристики коллекций персоналий

Характеристиками содержания коллекций персоналий являются свойства коллекции в целом и свойства отдельных персоналий. Под свойствами коллекции понимаются: критерий отбора лиц (хронологические, территориальные, профессиональные, социально-политические и пр. границы), число персоналий (масштаб), полнота заполнения коллекции по отношению к заявленным границам.

Содержательными характеристиками персоналий являются уровень детальности – наличие или отсутствие тех или иных аспектов биографии, а также их соотношенность с источниками (наличие точных указаний).

Если в ресурсе содержатся все персоналии, определяемые его границами, тогда пользователь либо найдет сведения об искомом лице, либо удостоверится, что это лицо не принадлежит к указанной группе лиц. Наличие указаний на источники позволяет как проверить данные, так и углубить представление об изучаемом индивидууме.

К сожалению, коллективный способ формирования коллекции, так много дающий в отношении её масштаба, сопряжен со стихийностью, неясностью критериев отбора, их случайностью и неполнотой по отношению к границам. Заметим, что качество содержания самой персоналии в случае, если коллективным ресурсом служит wiki, может быть и выше, чем в авторском ресурсе: благодаря открытому редактированию, сохранению версий.

6.2 Системные характеристики коллекций персоналий

Систематизацию также необходимо рассматривать на двух уровнях: коллекции в целом и отдельной персоналии.

Система – целостная совокупность элементов, взаимодействие которых обуславливает возникновение новых интегративных качеств, не присущих образующим ее частям, что отличает ее от простой суммы (конгломерата) объектов. Внутренняя организация системы, ее структура характеризуется формой связи ее компонентов и присущих им свойств.

С точки зрения коллекции персоналий системообразующее значение имеет фиксация связей как между персоналиями, так и между персоналиями и объектами другой природы, входящими в объемлющий коллекцию ресурс. Согласно работе [2], изучение биографии связано с выявлением «взаимосвязанных групп людей, составляющих жизнь данного человека: родственники, друзья, коллеги, враги, соседи по месту жительства, люди зависимые от поступков данного деятеля, наконец, все материальные и общественные сферы, опосредованно или непосредственно влияющие на данного человека».

Важнейшим механизмом упорядочения коллекции персоналий является типологизация или классификация [2] – ее разбиение на качественно определенные классы или типы на основе присущих этим объектам общих существенных, с точки зрения осуществляющего классификацию исследователя, признаков. Научная классификация требует существенных интеллектуальных затрат и редко воплощается в коллекциях персоналий. Вместо нее повсеместно используется формальная искусственная классификация по несущественным признакам: распределение по алфавитному порядку фамилий или по календарным датам рождений.

На уровне персоналии системность проявляется в ее внутренней структуре, в идеале представляющей многоступенчатое деление биографии по существенным признакам. Это также довольно сложная информационная работа: отдельные составляющие биографии могут быть связаны хронологически, логически, предметно. Очевидно, что хорошо структурированный информационный объект гораздо легче понять и интерпретировать, а также выбрать из него необходимое. Одним из видов структуризации является соотнесение данных из нескольких источников. В частности, в Хроносе используется механическое соединение текстов, благодаря чему сохранены разносторонние взгляды, но снижена наглядность, в Википедии произведена интеграция сведений, что повышает наглядность. В лучших статьях Википедии обозначены различия между источниками.

Заметим, что качество систематичности является важнейшим и для бумажных ресурсов справочно-информационного или научного содержания. Для электронных ресурсов оно играет роль базиса для подключения возможностей инструментария.

6.3 Инструментальные характеристики коллекций персоналий

Важнейшей инструментальной характеристикой коллекций персоналий является функциональность организации доступа к данным, в том числе, организация поиска/навигации, визуального представления. Простейший пример: если ресурс имеет простой, интуитивно понятный интерфейс, снижается время его освоения и требования к квалификации потребителя, что минимизирует суммарное время получения нужной информации.

Различные варианты поиска, используемые в инструментарии электронных коллекций, могут быть в той или иной степени востребованы при работе с коллекциями персоналий. Речь идет о полнотекстовом и атрибутом (многоаспектном) поиске в спектре возможных опций. Весьма полезны для биографического поиска (и весьма редко встречаются на практике) возможности атрибутного поиска по диапазонам дат и чисел. Наиболее эффективный и при этом чрезвычайно простой вариант поиска реализован в международной генеалогической базе familysearch. Поиск ведется по сочетанию (возможно неизвестных) имен-фамилий искомого лица, его родителей, детей, страны проживания, а, главное, времени указанного события: рождения, смерти, женитьбы, задаваемому диапазоном. Навигация (процесс перехода от одной страницы к другой, включающий промежуточные действия, в том числе, мыслительные операции в это время) в коллекциях персоналий, как правило, представляет специальный вариант поиска. Самый распространенный вид навигации – по уже упоминавшемуся формальному искусственному делению по алфавиту имен, а также году/дате рождения/смерти. В коллекциях, предусматривающих категоризацию (классификацию), возможна навигация по классификационным признакам (категориям, ключевым словам, тематике).

Чрезвычайно актуальна для коллекций персоналий навигация по связям между лицами: семейным, образовательным, профессиональным и пр. Родственные связи – основное содержание генеалогических коллекций. Не менее значима образовательная генеалогия – связь учитель – ученик, которую можно проследить в коллекциях «Математическая генеалогия» и «Биографии отечественных художников».

Эффективность навигации характеризуется числом шагов до искомого персоналии, а также вспомогательными интерфейсными приемами, облегчающими ориентировку.

Качество визуального представления ответа на поисковый запрос или очередного шага навигации определяется наглядностью и интерактивными возможностями. Очевидно, что плоский неформатированный текст менее нагляден, чем текст, использующий шрифты, выделение, рубрик таблицы. Высокой наглядностью обладают идеографические схемы. Такие схемы могут иллюстрировать связи между лицами; демографические, социальные, образовательные, профессиональные, общественно-политические и другие характеристики лица/лиц в зависимости от времени и во взаимных зависимостях; биографические события (например, на хронологической шкале, географической карте). Интерактивные возможности формы: гиперссылки, всплывающие окна, динамические объекты – важные аспекты в обеспечении эффективности доступа.

Разные исследования или разные аспекты одного исследования требуют разных форм представления данных. На языке информационных технологий

выбор формы представления связан с возможностями фильтрации, сортировки, группировки, определения визуальных характеристик. Ресурс, предоставляющий возможность выбора и настройки соответствующих параметров, очевидно, более применим, чем ресурс, их не предоставляющий.

Чрезвычайно важным инструментальным свойством коллекции персоналий является наличие в ней функций экспорта и импорта, посредством которых осуществляется перенесение данных из ресурса в ресурс, подключение автоматических средств анализа и обработки. Так же, как и другие области информационных технологий, эта сфера является полем стандартизации. Эти вопросы требуют отдельного обсуждения.

7 Заключение

В работе представлены основные виды электронных биографических ресурсов, перечислены типовые задачи биографического поиска, рассмотрены важнейшие характеристики биографических ресурсов, влияющие на его эффективность. Главным результатом нашего рассмотрения является представление о коллекции персоналий, как о систематизированном, в той или иной степени, наборе персональных биографических данных, допускающих спектр возможных применений. При этом выявлена общность справочно-информационных, просопографических и генеалогических ресурсов.

В ближайшие задачи дальнейших исследований входят уточнение характеристик и показателей качества биографических ресурсов и их оценка для представительного ряда биографических ресурсов. В результате должна быть сформирована матрица, соотносящая успешность решения различных задач биографических исследований с характеристиками биографических ресурсов, а также выявлены примеры удачных конструктивных решений.

Конечную цель работ можно было бы сформулировать как информационно-методическое обеспечение «средств реконструкции причинно-следственных связей и конструирования личных досье на основе анализа всей совокупности имеющейся о личности информации во всемирной информационной сети» – одной из глобальных задач исторической информатики по [2]. Однако в более рациональной перспективе речь может идти о методах и средствах частичного приближения к этому результату. В частности, о методах и средствах организации доступа к коллекциям персоналий, интеграции персональных данных из разнородных источников, рациональной организации информационного пространства биографического исследования.

Интерес как профессиональных исследователей, так и любителей к культурно-историческому феномену – биографии растет, стремительно растет и объем данных, доступных в электронном виде. Тем важнее и актуальнее является перечисленный комплекс задач.

Литература

- [1] Бекжанова Н.В. Метод библиографической реконструкции биографии и его использование для справочно-библиографического обслуживания // Библиосфера. – 2006. – С. 5-11.
- [2] Бочаров А.В. Алгоритмы использования основных научных методов в конкретно-историческом исследовании. Учебное пособие. – Томск, 2007.
- [3] Валевский В.Л. Биографика как дисциплина гуманитарного цикла // Лица: биографический альманах. – СПб.: Феникс, 1995. – Вып. 6. – С. 33-68.
- [4] Генеалогическая информация в государственных архивах России: справочное пособие / Росархив, ВНИИДАД. – М., 2004. – 280 с.
- [5] Диковская Е.А. Эволюция персонально-ориентированных элементов справочно-поискового аппарата современных библиотек: Автореф. дис. ... канд. пед. наук; Сам. гос. акад. культуры и искусств. – Самара, 2004. – 20 с.
- [6] Концепция информатизации архивного дела России / Утверждена Росархивом в 1995 г. – <http://www.rusarchives.ru/informatization/conseption.shtml>.
- [7] Перевертень В.А. Модели организации информации в информационных системах для просопографических исследований // Круг идей: традиции и тенденции исторической информатики. Труды IV конф. Ассоциации «История и компьютер». – М., 1997. – С. 42-58.
- [8] Петриченко М.Б. Компьютер в генеалогических исследованиях. – М.: РОИА, 2004. – 365 с.
- [9] Руководство UNIMARC / ГПНТБ России I. – М., 1992.
- [10] Сидорова В.А. и др. Избранная библиография по отечественной генеалогии // Вестник архивиста. – 1997. – № 4 (40). – С. 33-61; № 5 (41). – С. 98-123; 1998. – № 1 (43). – С. 91-105.
- [11] Сукиасян Р.Э. Классификационный (систематический) поиск в эпоху электронных каталогов (статья вторая) // Научные и технические библиотеки. – 2009. – № 10. – С. 38-47.
- [12] Юмашева Ю.Ю. Историография просопографии // Изв. Уральского государственного университета. Гуманитарные науки. 2005. – Вып. 10, № 39. – С. 95-127.

Digital biographic resources

N. Markova, I. Adamovich

The paper considers biographical information collections. It provides a review of their current state, problems and perspectives. Scenarios of biographical research are considering. The research efficiency depends on collection in use characteristics. The paper defines characteristics and gives their analysis.