



МОДЕЛИ ДАННЫХ ДЛЯ ВИРТУАЛЬНОЙ ИСТОРИЧЕСКОЙ РЕКОНСТРУКЦИИ МАЛЫХ СОЦИАЛЬНЫХ ГРУПП НА ОСНОВЕ АРХИВНЫХ ДОКУМЕНТОВ XVII – НАЧАЛА XX ВЕКА

Основной целью исследования является создание моделей данных для системы исторической реконструкции малых социальных групп на основе архивных документов XVII – начала XX века по Вологодскому уезду. Определен круг регулярных источников информации и построены модели данных этих источников с учетом их исторической изменчивости, выявлены закономерности связей и возможности их идентификации по нечетким и косвенным данным. Приведены результаты анализа средств обработки и представления генеалогической информации и требования к функциональности системы для исторической реконструкции и аналитических исследований.

Архивный документ, генеалогическая информация, идентификация родственных связей, виртуальная историческая реконструкция, модели данных, нечеткие данные, функциональные требования, экспорт и импорт данных.

Современные информационные технологии предоставляют исторической науке все больше возможностей для реконструкции характеристик среды обитания и условий существования наших предков. Огромные объемы информации, хранящейся в архивах, постепенно становятся доступными для компьютерной обработки. При наличии и доступности источников можно, например, построить схему последовательности N поколений прямых предков для любой персоны, насчитывающую $2 \cdot N - 1$ имен (задача генеалогии), или представить картину событий хозяйственной и общественной жизни на территории в определенный период прошлого (задача исторической реконструкции). Реконструкция фактов и явлений прошлого принципиально отличается от моделирования будущего тем, что мы достоверно наблюдаем их конечный результат. Проблема лишь в том, что информация для построения реконструкций носит фрагментарный и нечеткий характер, а доступ к ней существенным образом ограничен.

Для преодоления пробелов и нечеткостей информации полезно пользоваться моделями процессов и явлений прошлого, которые возможно построить на основе изучения и использования известных и вновь открываемых в ходе исторического исследования закономерностей. Несомненную ценность для изучения прошлого могут представлять модели данных малых социальных групп, таких как семья, род, отдельное поселение, община, в первую очередь генеалогической информации.

В настоящее время сбор и обработка первичной информации о малых социальных группах ведется вручную или с использованием текстовых и табличных процессоров, реже – универсальных или специализированных СУБД. Главная проблема при этом – трудоемкость распознавания и ввода первичной информации, что ведет к неполноте и искажениям данных, а также ограниченный набор первоначально используемых атрибутов объектов, который при развитии исследования по необходимости приходится периодически расширять.

Актуальность настоящей работы обусловлена несовершенством и низкой производительностью существующих технологий поиска, ограниченной доступностью, неполнотой и нечеткостью данных архивных документов. Работа посвящена анализу доступных архивных источников информации и технологий ее поиска и обработки, формированию требований к функциональности системы для исторической реконструкции и аналитических исследований характеристик и связей в малых социальных группах на основе неполных и нечетких данных.

Базовой технологией получения первичной для построения моделей информации является работа с первоисточниками в читальных залах архивов. Для поиска информации используется развитый нормативно-справочный аппарат и различные методические рекомендации [1]. Копирование информации из архивных документов возможно путем составления выписок, фотографии отдельных документов или их фрагментов с последующим распознаванием и фиксацией в бумажном и электронном виде.

Помимо возможности работы в читальном зале архива, часть документов по Вологодскому уезду доступна на интернет-портале архивной службы Вологодской области [2] как в виде сканов документов, так и достаточно подробной метаинформации по ним. Возможен поиск документов по описям и частично доступ к их сканам в электронном виде. Там же ведется периодически пополняемый Государственный Реестр уникальных документов, содержащий управленческую, хозяйственную и генеалогическую информацию периода XVI – начала XX веков. В сети Интернет реализуется проект «Вернем нашим предкам имена» [3], целью которого является создание на основе исповедных ведомостей электронной базы данных по жителям Вологодского уезда начала XX века. Развиваются поисковые системы типа Базы знаний портала «familio» [4], активно функционируют различные генеалогические форумы, где идет обмен информацией по источникам, персонам и технологиям поиска [5].

Решению задачи поиска информации в архивных источниках, ее распознавания и интерпретации посвящено немало исследований, результаты которых отражены в диссертационных работах [6, 7], методических рекомендациях, научных публикациях, учебной и популярной литературе [8]. В них приводятся подробные характеристики источников с точки зрения наборов доступных данных, возможных интерпретаций описываемых фактов, форм и содержания производных документов. С другой стороны, в мире динамично развиваются идеи и технологии исторической реконструкции различных событий, объектов, процессов и явлений прошлого, формируются научные и методические основы этого направления [9, 10], реализуются программы дополнительного образования. К сожалению, по большей части эти исследования касаются военно-исторической и военно-политической тематики, моделирования архитектурных объектов, предметов вооружений или макросистем, а не особенностей мирной жизни и быта наших предков на уровне семьи, рода, поселения.

Комплексной методики, предусматривающей возможность использования прямых и косвенных данных для верификации полученной в ходе архивного поиска информации, устранения неполноты и нечеткости данных, интерпретации фактов и выявления скрытых зависимостей в предметной области исследования, пока не предложено.

Программные продукты и сервисы для генеалогических исследований, доступные на рынке, в большинстве своем предоставляют лишь возможность ввода предварительно найденной, отобранной и выверенной информации в специальные формы (карточки) и построения выходных форм в виде различных графических схем, списков и таблиц [11, 12]. Для создания виртуальных исторических реконструкций используются преимущественно универсальные инструменты 2D- и 3D-моделирования или игровые движки и платформы разработки типа Unity3D [13].

Цель работы: на основе анализа структуры регулярных архивных источников информации создать модели данных для системы исторической реконструкции малых социальных групп на основе архивных документов XVII – начала XX веков по Вологодскому уезду и определить требования к функциональности системы, позволяющей обрабатывать нечеткие данные.

Объект исследования: источники статистической, управленческой, экономической, демографической и генеалогической информации XVII – начала XX веков среди архивных документов по Вологодскому уезду и технологии их анализа.

В качестве основных методов исследования были использованы как исторические методы (историко-системный, источниковедческий, антропонимический анализ), так и общенаучные (анализ, синтез, аналогии, дедукция), а также методы и технологии исторической информатики, системного моделирования.

Задачи исследования:

1. Определение круга и характеристик доступных регулярных источников информации о малых социальных группах в Вологодском уезде в исторической ретроспективе XVII – начала XX веков.

2. Построение моделей данных основных источников информации (метрическая книга, ревизская

сказка, исповедная ведомость), позволяющих вводить и обрабатывать нечеткие данные, и формирование основных справочников модели.

3. Выявление основных закономерностей связей в малых социальных группах и возможностей их идентификации по данным архивных документов.

4. Анализ доступных программных средств обработки и представления генеалогической информации для целей исторической реконструкции.

5. Разработка функциональных требований к информационно-системе для исторической реконструкции и аналитических исследований.

Научная новизна работы заключается в синтезе расширенных за счет косвенных признаков и отношений (родства, свойства, территориальных и духовных связей) моделей регулярных источников данных, допускающих ввод и обработку нечеткой информации.

Практическая значимость работы заключается в определении функциональных требований к информационной системе для исторической реконструкции и аналитических исследований малых социальных групп, позволяющей вводить и обрабатывать нечеткие данные.

Представим кратко основные результаты работы.

Первый результат. Определен круг и характеристики доступных регулярных источников информации о малых социальных группах в Вологодском уезде в исторической ретроспективе XVII – начала XX веков. Под регулярными источниками будем понимать виды документов, составленных по определенной форме, одинаковой для различных социальных групп на всей территории уезда, и содержащих информацию, фиксируемую в течение гораздо большего по сравнению с заданной периодичностью сбора времени. К таким видам источников относятся переписная книга, ревизская сказка, метрическая книга, исповедная ведомость, брачный обыск и ряд других. Для представителей духовенства к числу регулярных источников можно добавить клировые ведомости, которые, в силу относительно небольшой доли духовенства в общей численности населения, детально не рассматривались в проекте. В целом же для исследования малых социальных групп могут быть использованы более 120 различных видов архивных документов [1].

В течение исследуемого периода времени каждый из указанных видов документов неоднократно претерпевал изменения в составе атрибутов описываемых объектов и событий как в сторону увеличения их количества, степени детализации и правил оформления (например, исповедная ведомость), так и в обратном направлении (например, ревизская сказка). Территориальная принадлежность малой социальной группы в светских документах определялась по правилам административного деления (губерния – уезд – треть – волость – землевладелец – населенный пункт), в церковных: епархия – уезд – церковный приход – населенный пункт – землевладелец, причем один и тот же населенный пункт мог быть разделен не только между землевладельцами, но и между приходами. В изучаемом периоде территориальные границы объектов административного деления, названия ряда населенных пунктов и их принадлежность неоднократно менялись, что требует соответствующего учета в атрибутах описания и правилах анализа характеристик семей и иных малых социальных групп.

Практика исследований показывает, что уже на первом этапе поиска полезно вводить в базу совокупность данных не только по искомым семьям (дворам), а в целом по деревне или социальной группе (землевладельцу), благо в уезде они в большинстве своем насчитывали до одного – двух десятков дворов. Это обстоятельство объясняется господством в исследуемом периоде натурального хозяйства и чрезвычайно малой мобильностью основных групп населения, за исключением дворян и духовенства.

Основные атрибуты записей в переписной книге: имя и отчество (иногда – прозвище/фамилия) владельца двора, имена и отчества взрослых мужчин как возможных преемников тягла с указанием родственных связей и рода занятий – источника существования и сбора налогов. В документах второй половины XVII века фиксировался и возраст несовершеннолетних детей.

Основные атрибуты записей в метрической книге.

Рождения: дата рождения, имя новорожденного, статус семейный, место жительства, землевладелец; статус социальный, фамилия, имя и отчество отца (для умершего – обязательная отметка); статус семейный (жена, вдова), имя и отчество матери; статус социальный, статус семейный по отношению к отцу или новорожденному, фамилия, имя и отчество, место жительства восприемника.

Браки: дата венчания; статус, имя, отчество, фамилия жениха, место жительства, землевладелец, возраст, номер брака; статус, фамилия, имя и отчество отца жениха; статус, имя, отчество, фамилия невесты, происхождение, землевладелец, возраст, номер брака; статус, фамилия, имя и отчество отца невесты; роль, статус социальный, статус семейный, фамилия, имя и отчество поручителя, место жительства поручителя.

Смерти: дата; место жительства, статус социальный и семейный, имя, отчество, фамилия человека, возраст, причина смерти, место захоронения.

Основные атрибуты записей в исповедной ведомости.

Год составления, уезд, треть, приход, место жительства, землевладелец, двор (№ записи), статус социальный, статус семейный, имя, отчество, фамилия человека, возраст, присутствие на исповеди.

Основные атрибуты записей в ревизской сказке.

Уезд, волость, место жительства, землевладелец, двор, для вновь прибывших – основание и адрес выбытия, статус социальный, статус семейный, имя, отчество, фамилия человека, возраст по предыдущей и текущей ревизии, год выбытия/перевода, причина выбытия, землевладелец, место выбытия. Для замужних женщин/вдов дополнительно – откуда взята в замужество, наличие документа-согласия.

В брачном обыске набор атрибутов персон практически тот же, что и в разделе «браки» метрической книги, но сопровождается отметкой о наличии согласия родителей или владельцев (для крепостных), автографами жениха, невесты и поручителей или уполномоченных ими лиц.

Фиксация прочих атрибутов записей (например, даты и участники церковных треб) для целей исторической реконструкции ценности не имеет.

Дополнительный признак каждой записи или их группы – источник информации (архив, фонд, опись, дело, лист).

Отметим, что даже прекрасно читаемый почерк писца довольно сложно распознать с помощью современных программных средств: затраты на обучение системы распознавания оказываются больше, чем на ввод информации по отдельным объектам вручную.

Второй результат. Построены модели данных основных источников информации (метрическая книга, ревизская сказка, исповедная ведомость), позволяющих вводить и обрабатывать нечеткие данные, сформированы основные справочники модели данных.

Рассмотрим пример фрагмента записи в исповедной ведомости Троицкого в Подлесном прихода за 1825 год по Ямской Московской Слободы д. Головиной ямщикам (ГАВО, ф.496, оп.19, д. 280, л. 494об.):

«...14. Андрей Матвеев, 68, жена его Катерина Иванова, 51, дети их Иван, 43, Матрена, 33, Настасья, 21, Иванова жена Евдокия, 43, дети их Александр, 13, Василий, 11, Анна, 6...»

Фрагмент отображения этой записи в базу данных представлен на рисунке. Здесь даты рождения и смерти являются нечеткими данными и указаны в диапазонах значений, написание имен приведено к современному виду по справочнику, а косвенные данные формируются в позициях «свекор» и «тесть».

Основные источники нечеткостей и неполноты данных в архивных документах: физический износ и обрезка при переплете (особенно у сшитых метрических книг), неразборчивость текста и ошибки писца, антропонимические подстановки. В документах XVII–XVIII веков встречалось, например, что у замужних женщин-крестьянок вместо имени отца писалось имя мужа, а для мужчин, «вышедших в дом», иногда использовалась фамилия тестя, пока тот был главой семьи. При вводе и обработке нечетких данных возможны два варианта действий: множественные значения нечетких полей или использование нечетких правил в поисковых запросах, что обычно полезно уже на стадии анализа при попытке преодоления неполноты данных. Кроме вариантов символического и числового (дата, возраст) представления как результата распознавания рукописного текста в случае возникновения нечеткости полезно поле с графическим представлением фрагмента записи.

Для автоматизации ввода информации, уменьшения потока ошибок и нечеткостей сформированы справочники модели данных: имена, социальные и семейные статусы субъектов, населенные пункты, единицы административно-территориального деления, землевладельцы. В справочнике имен приводятся варианты современного и исторического написания с возможными синонимами. Примеры семейного статуса: незаконнорожденный, двойня, вдова, брат, сестра и т.п. В справочнике населенных пунктов учтены смены названий в исторической ретроспективе (по интервалам времени действия названия). Справочник землевладельцев содержит их титулы, статус, фамилии, имена и отчества. При вводе данных по населенному пункту приводится интервал времени владения объектом и причина перехода права (приданое, дарение, купля-продажа, наследование, конфискация).

Год	Приход	Деревня	Двор	Идент.	Фамилия	Имя	Имя отца	Дата рожд.	Дата смерти	Статус социал.	Статус семейн.
	-	-	-	0	-	Матвей	-	Не позднее 1741	Не позднее 1825	-	свекор
1825	Троицкий в Подлесном	Головино	14	0/0	-	Андрей	Матвей	1756-1757	Не ранее 1825	ямщик	муж
	-	-	-	1	-	Иван	-	Не позднее 1758	-	-	тесть
1825	Троицкий в Подлесном	Головино	14	1/0	-	Екатерина	Иван	1773-1774	Не ранее 1825	ямщик	жена
1825	Троицкий в Подлесном	Головино	14	0/0/1	-	Иван	Андрей	1781-1782	Не ранее 1825	ямщик	сын
1825	Троицкий в Подлесном	Головино	14	0/0/2	-	Матрена	Андрей	1791-1792	Не ранее 1825	ямщик	дочь
1825	Троицкий в Подлесном	Головино	14	0/0/3	-	Анастасия	Андрей	1803-1804	Не ранее 1825	ямщик	дочь
1825	Троицкий в Подлесном	Головино	14	2/0/0	-	Евдокия	-	1781-1782	Не ранее 1825	ямщик	невестка
1825	Троицкий в Подлесном	Головино	14	0/0/0/0	-	Александр	Иван	1811-1812	Не ранее 1825	ямщик	внук
1825	Троицкий в Подлесном	Головино	14	0/0/0/1	-	Василий	Иван	1813-1814	Не ранее 1825	ямщик	внук
1825	Троицкий в Подлесном	Головино	14	0/0/0/2	-	Анна	Иван	1818-1819	Не ранее 1825	ямщик	внучка

Рис. Пример отображения записи исповедной ведомости в базу данных

Третий результат. Выявлены основные закономерности связей в малых социальных группах и возможности их идентификации по данным архивных документов. Наряду со стандартными и, как правило, четкими связями отношений родства или свойства, выявлены традиции, когда восприемники сыновей в семье брались со стороны матери, а дочерей – со стороны отца, или поручители на венчании: первый со стороны отца, второй – матери, но по разным причинам могло быть и иначе (например, инверсная традиция при «выходе» жениха в дом невесты). Даже расположение семей в соседних номерах дворов поселения в исповедной ведомости часто указывало на семейное родство (взрослые сыновья выделились в отдельные дворы). Анализ данных в приведенном на рисунке примере на соотношение возрастов Ивана Андреева и Екатерины Ивановой говорит о том, что она – мачеха Ивану Андрееву и не первая жена Андрея Матвеева, но факт требует подтверждения метриками или исповедными ведомостями более ранних лет. Это – примеры косвенных признаков, полезных для поиска при недоступности прямых подтверждающих документов.

Четвертый результат. Анализ доступных программных средств обработки и представления генеалогической информации показал, что они в большинстве своем дают лишь возможность ввода предварительно найденной, отобранной и выверенной информации в специальные формы (карточки) и построения выходных форм в виде различных графических схем, списков и таблиц [11, 12], в лучшем случае лишь с формальным контролем родственных связей по возрастам и сочетаний имен. Представление нечетких данных в них возможно также преимущественно только относительно дат и возрастов, например «(не)ранее», «(не)позднее», «между» и т.п.

Исследование программных средств, применяемых для виртуальных исторических реконструкций, показало, что это в основном 2D-редакторы (графические редакторы типа Adobe Photoshop, Corel Draw и проч.), 3D-

редакторы (AutoCAD, Autodesk 3D Max и проч.) и «движки» 3D engines, среди которых наиболее популярна платформа для работы с интерактивным RT3D-контентом Unity3D. Однако все они имеют универсальный характер и требуют для исторической реконструкции создания специальных библиотек элементов и утилит, например привязки к реальному географическому пространству. Возможно в перспективе и использование свободно распространяемых ГИС-платформ.

Пятый результат. Разработаны функциональные требования к информационной системе для исторической реконструкции и аналитических исследований. Среди этих требований:

- символьный ввод данных по структурам регулярных документов (ревизская сказка, метрическая книга, исповедная ведомость) с использованием множественных значений для нечетких данных, справочников и наследования значений полей в последовательных записях;
- ввод графических образов фрагментов наиболее важных документов (метрические записи, автографы, портреты, фотографии);
- создание карточек персон и населенных пунктов с набором символьных, числовых и графических атрибутов, связей и текстовых комментариев для нерегулярной информации;
- использование фильтров при поиске и отображении информации;
- отображение событий жизненного цикла персоны или малой социальной группы с характеристиками времени и места;
- вычисление статистических характеристик по малым социальным группам;
- вывод на экран или на печать персональных карточек, таблиц, генеалогических деревьев, родственных росписей по поколениям или на определенную дату;
- возможность сортировки информации в таблицах по значениям отдельных полей;
- экспорт и импорт данных формата Gedcom;

- экспорт и импорт изображений в растровом и векторном форматах;

- создание резервной копии данных, возможность объединения, разделения и синхронизации данных на нескольких рабочих местах.

В числе нефункциональных требований:

- кроссплатформенность (Windows7 и выше, клоны Unix, в т.ч. Android 4 и выше);

- использование свободно распространяемых сред разработки, утилит, библиотек, как вариант, Unity3D или ГИС-платформы;

- документирование кода и сопутствующих данных для обеспечения независимой поддержки и развития продукта.

Выводы. В ходе исследования проведен анализ структуры регулярных источников генеалогической информации и доступных технологий ее поиска и обработки, предложены модели данных регулярных источников, допускающие ввод нечеткой информации, выявлены возможности идентификации социальных связей по нечетким данным, разработаны требования к информационной системе для исторической реконструкции и аналитических исследований. На следующих этапах работ планируется создание логической схемы данных системы, проектирование ее бизнес-логики и пользовательского интерфейса, реализация системы в среде Unity3D или свободно распространяемых ГИС-платформ.

Литература

1. Генеалогическая информация в государственных архивах России : справочное пособие / С. Н. Романова, И. И. Глуховская, М. П. Дьячкова, В. И. Звачич. – Москва : Федеральное архивное агентство, ВНИИДАД, 2004. – 280 с. – Текст : непосредственный.

2. Интернет-портал архивной службы Вологодской области : [сайт]. – URL: <https://gosarchive.gov35.ru/> (дата обращения: 17.09.2022). – Текст : электронный.

3. Вернем нашим предкам имена : [сайт]. – URL: <https://vk.com/vernemimena> (дата обращения: 17.09.2022). – Текст : электронный.

4. База знаний портала familio : [сайт]. – URL: <https://familio.org/knowledge-base> (дата обращения: 17.09.2022). – Текст : электронный.

5. Генеалогический форум ВГД : [сайт]. – URL: <https://forum.vgd.ru/> (дата обращения: 17.09.2022). – Текст : электронный.

6. Ильина, Е. В. Источники и методы восстановления генеалогии крестьянских родов середины XVIII –

начала XX в.: на материалах Алтайского (Кольваново-Воскресенского) горного округа : специальность 07.00.09 «Историография, источниковедение и методы исторического исследования» : автореферат диссертации на соискание ученой степени кандидата исторических наук / Ильина Елена Викторовна ; Алтайский государственный университет. – Барнаул, 2010. – 23 с. – Текст : непосредственный.

7. Пашинин, А. В. Статистические источники по генеалогии крестьянских родов (семей) конца XVII – начала XX вв. в фондах государственного архива республики Бурятия : специальность 07.00.09 «Историография, источниковедение и методы исторического исследования» : автореферат диссертации на соискание ученой степени кандидата исторических наук / Пашинин Александр Васильевич ; Институт монголоведения, буддологии и тибетологии СО РАН. – Улан-Удэ, 2015. – 24 с. – Текст : непосредственный.

8. Савелов, Л. М. Лекции по русской генеалогии / Л. М. Савелов. – Москва : Археографический центр, 1994. – 302 с. – Текст : непосредственный.

9. Побережников, И. В. Пространственно-временная модель в исторических реконструкциях модернизации : специальность 07.00.09 «Историография, источниковедение и методы исторического исследования» : диссертация на соискание ученой степени доктора исторических наук / Побережников Игорь Васильевич ; Институт истории и археологии УрО РАН. – Екатеринбург, 2011. – 550 с. – Текст : непосредственный.

10. Митюков, Н. В. Математические модели и программные средства для реконструкции военно-исторических данных : специальность 05.13.18 «Математическое моделирование, численные методы и комплексы программ» : автореферат диссертации на соискание ученой степени доктора технических наук / Митюков, Николай Витальевич ; Российский новый университет. – Ижевск, 2009. – 35 с. – Текст : непосредственный.

11. Древо Жизни : [сайт]. – URL: <https://genery.com/ru/> (дата обращения: 17.09.2022). – Текст : электронный.

12. MyHeritage : [сайт]. – URL: <https://www.myheritage.com/?lang=RU> (дата обращения: 17.09.2022). – Текст : электронный.

13. Платформа для работы с интерактивным RT3D-контентом Unity3D : [сайт]. – URL: <https://unity.com/ru> (дата обращения: 17.09.2022). – Текст : электронный.

A.M. Polyansky
Vologda State University

DATA MODELS FOR VIRTUAL HISTORICAL RECONSTRUCTION OF SMALL SOCIAL GROUPS BASED ON ARCHIVAL DOCUMENTS OF THE XVII – EARLY XX CENTURY

The main purpose of the study is to create data models for the system of historical reconstruction of small social groups based on archival documents of the XVII – early XX centuries in the Vologda uyezd. The range of regular sources of information is determined and data models of these sources are constructed taking into account their historical variability, patterns of connections and the possibility of their identification by fuzzy and indirect data are revealed. The results of the analysis of the means of processing and presenting genealogical information and the requirements for the functionality of the system for historical reconstruction and analytical research are presented.

Archival document, genealogical information, identification of family ties, virtual historical reconstruction, data models, fuzzy data, functional requirements, data export and import.