

СИСТЕМАТИЗАЦИЯ И СОХРАНЕНИЕ ИСТОРИЧЕСКИХ ИСТОЧНИКОВ В УСЛОВИЯХ ЦИФРОВОЙ ЭПОХИ

Т.Е. Болдовская^{1,2}

к.т.н., доцент, e-mail: boldovskaya73@gmail.com

М.И. Ремизов^{1,3}

младший научный сотрудник, e-mail: remizov.maximus.maxim6@gmail.com

А.Е. Ветров^{1,3}

младший научный сотрудник, e-mail: aleksandrvetrv838@gmail.com

¹Омский государственный технический университет, Омск, Россия

²Омский государственный университет им. Ф.М. Достоевского, Омск, Россия

³Томский государственный университет, Томск, Россия

Аннотация. Статья посвящена проблеме сохранения исторического наследия через цифровизацию письменных источников. Рассматриваются существующие электронные хранилища и платформы, а также процесс разработки информационно-аналитического онлайн-ресурса «Православный ландшафт таёжной Сибири». Основное внимание уделяется унификации данных, созданию многоуровневой системы хранения и проектированию баз данных.

Ключевые слова: исторические источники, база данных, унификация данных, электронные архивы, цифровая гуманитаристика, цифровое наследие, ГИС-технологии, межконфессиональные связи.

Введение

Существует множество источников исторической информации, одним из них являются письменные материалы, содержащие информацию о культурном наследии мировой истории, благодаря им можно восстановить наиболее полную историческую картину прошлого. К сожалению, подлинники исторических документов в процессе естественного старения, ненадлежащего хранения и обращения могут быть утрачены, что несёт огромный урон для культурного наследия. Для борьбы с этой проблемой разрабатываются различные электронные хранилища данных, но их главным недостатком является нестабильность. Программы и платформы сменяются так быстро, что невозможно организовать надёжную сохранность данных, ведь уже через несколько лет платформа будет попросту недоступна из-за несовместимости с новой системой [1]. В свою очередь, оцифрованные экземпляры исторических документов не имеют единого стандарта, возникает риск несоответствия копии оригиналу, что заметно усложняет работу с историческими данными [2]. Информационно-аналитический онлайн-ресурс «Православный ландшафт таёжной Сибири» разрабатывается для решения этих проблем. Текущая цель проекта – отработать применение цифровых технологий на ограниченной территории, чтобы в

дальнейшем их можно было масштабировать и использовать для решения более глобальных задач в области сохранения и исследования исторического наследия. Все исторические данные, загруженные в нашу систему, будут оцифрованы и приведены к единому стандарту, что ускоряет исторические исследования и повышает удобство при работе с данными, ведь пользователь сможет просматривать данные в удобном для него формате, например, в текстовой или в графической форме. Это также позволит сократить использование непосредственно самих письменных экземпляров, ведь теперь появится возможность работать одновременно с множеством источников в одном месте, это способствует более длительному сохранению физических оригиналов в хорошем состоянии. Кроме того, благодаря многоуровневой системе хранения цифровых копий ценных исторических данных обеспечивается их долговременная сохранность и доступность, минимизируя вероятность необратимой потери информации.

1. Примеры других платформ

На сегодняшний день существует множество веб-платформ, которые нацелены на создание цифровых архивов. Например: «Europeana» – проект, основной целью которого является создание цифровой библиотеки, музея и архива Европы. Он был запущен в 2008 г. Европейской комиссией и является коллективным усилием более 3500 учреждений, включая музеи, библиотеки, архивы и галереи из разных стран Европы. Цель проекта «Europeana» – собрать, сохранить и сделать доступными цифровые копии культурных объектов и артефактов, связанных с историей, культурой и наследием Европы [3].

«Mapping the Republic of Letters» – масштабная платформа, которая стремится воссоздать и визуализировать сети общения между интеллектуальными деятелями Европы в период с XVI по XVIII в. Целью проекта является представление исторического контекста, в котором развивались и распространялись идеи, знания и культурные ценности этого времени. Проект основан на обширном анализе письменной переписки, архивных документов и других исторических источников. Исследователи проекта изучают и систематизируют сотни тысяч писем, которые были обменены между известными мыслителями, учёными, писателями и философами того периода. Они анализируют содержание писем, контексты общения и сети связей, чтобы понять, как эти интеллектуальные деятели взаимодействовали друг с другом и как их идеи распространялись по всей Европе. Визуализация этих сетей общения и обмена знаниями является ключевым аспектом проекта «Mapping the Republic of Letters». С использованием географических и социальных данных исследователи создают интерактивные карты и графические представления, которые позволяют исследователям и публике лучше понять масштаб и значимость обмена идеями в этом периоде истории. Эти визуализации помогают выявить основные центры интеллектуальной активности, показать связи между различными деятелями и организациями, а также иллюстрировать потоки идей и знаний [4].

«PRO Сибирь» – онлайн-сервис, который в настоящее время предлагает доступ к более 120 тысячам источников, включая периодические издания, рукописи, книжные памятники, материалы исследовательских экспедиций, архивы ученых и дру-

гие ресурсы, отражающие важность Сибири и Дальнего Востока с точки зрения их территориального, экономического и культурно-исторического значения. Главной особенностью платформы «PRO Сибирь» является её интерактивный характер, который позволяет пользователям принимать активное участие в её развитии благодаря интеграции поисковых, аналитических и коммуникационных сервисов. На платформе пользователи могут выступать в роли исследователей и волонтеров, а также использовать различные функциональные возможности. Они могут распознавать тексты, размещать свои собственные оцифрованные материалы, описывать и тегировать информацию, создавать тематические коллекции, публиковать статьи в блогах, находить единомышленников и предлагать новые идеи. Таким образом, платформа предоставляет возможность активного взаимодействия пользователей, содействуя развитию и обогащению её содержимого [5].

Анализируя работу данных платформ, было принято решение разработать веб-сервис, который представлял бы собой некие исторические данные, экстрагированные из подлинников исторических документов и представленные на современной карте. При этом карта будет иметь поддержку графов, фильтров и слоев для более детального анализа.

2. Проектирование базы данных

Разработка базы данных для исторических исследований представляет собой сложный и многогранный процесс, требующий как технического проектирования, так и глубокого понимания предметной области. Обработка исторических данных связана с определёнными специфическими трудностями. Высокая степень взаимосвязанности объектов усложняет процесс их структурирования, поскольку один и тот же объект может присутствовать в разных исторических контекстах, различающихся во времени, пространстве и тематической направленности. Особое внимание было уделено точной репрезентации межобъектных связей, а также стандартизации терминологического аппарата, что позволило исключить дублирование и гарантировать высокую точность обработки данных. Важнейшей задачей стало создание системы, способной обрабатывать большие объёмы информации и гибко настраиваться под разнообразные исследовательские запросы. В этой связи критически важным стал выбор оптимальной методологии проектирования базы данных, обеспечивающей надёжное хранение информации наряду с удобным доступом для аналитических операций и визуализационных процедур.

Основываясь на проведённом анализе, было принято решение применить нисходящую методологию разработки базы данных ввиду её большей применимости к сложным информационным системам с множеством атрибутов, где идентификация всех функциональных зависимостей является сложной задачей. В противоположность восходящей методике, удобной для создания менее сложных баз данных с ограниченными наборами атрибутов, нисходящая стратегия позволяет глубже проработать семантику предметной области. Основные достоинства нисходящего метода заключаются в минимизации рисков возникновения ошибок в процессе эксплуатации информационной системы, высокой степени формализованности проекторочных процедур и снижении трудовых затрат на нормализацию даталогической

модели до необходимой нормальной формы. Процесс нисходящего проектирования состоит из пяти последовательных стадий [6]:

1. Анализ предметной области. На данном этапе осуществляется всесторонний анализ предметной области с последующей декомпозицией на отдельные элементы. В результате получается описание внешнего уровня базы данных, которая создаётся для пользователей и абстрагируется от особенностей реализации.
2. Разработка информационно-логической модели предметной области. Информационно-логическая модель отражает предметную область в виде совокупности информационных объектов и их структурных связей. Её разработка начинается с построения инфологической модели, затем на её основе строится концептуальная модель данных.
3. Формирование даталогической модели базы данных. Даталогическая модель базы данных отражает логические взаимосвязи между элементами данных, независимо от их содержания и физической организации. Она разрабатывается на основе информационно-логической модели предметной области с учётом конкретной реализации системы управления базой данных и специфики конкретной предметной области.
4. Этап нормализации. Здесь осуществляется процесс нормализации, который предполагает последовательное создание ряда моделей и применение нисходящего подхода к проектированию с минимальными корректировками. В качестве основы для нормализации используется даталогическая модель данных, разработанная на предшествующем этапе.
5. Формирование физической модели базы данных. На этом этапе осуществляется разработка физической модели базы данных, которая включает определение местоположения данных, методов доступа к ним, а также выбор техники индексирования. Данная модель также известна как внутренняя модель системы.

Концептуальное проектирование представляет из себя основу всего процесса проектирования баз данных. Для того чтобы база данных адекватно отражала предметную область, нужно было хорошо представлять себе все особенности, присущие данной предметной области, и уметь отобразить их в базе данных. Поэтому перед началом проектирования базы данных необходимо было детально изучить предоставленную нам историческую информацию, территориально-административное деление российской империи, территориально-административную иерархию Русской православной церкви, чтобы хорошо представлять себе все особенности, присущие данной предметной области. На данном этапе значительное содействие было оказано коллегами из Томского государственного университета. В рамках серии организованных онлайн-конференций были проведены детальные обсуждения и представлены исчерпывающие ответы на возникающие исследовательские и технические вопросы. Инфологическая модель предметной области может описываться как аналитическими, так и графическими средствами. Графическое представление является наиболее наглядным и простым для восприятия и анализа. Кроме того, графическое представление поддерживается системами автоматизации проектирования БД. Поэтому для описания объектов и их связей мы воспользовались графиче-

ским способом отображения, позволяющим построить модель, условно называемую «объект – свойство – связь». При исследовании и анализе предметной области были выделены следующие классы объектов: населённый пункт, церковь, образование, причт, население, социальные конфликты, религиозные девиации. Были отражены связи между объектом и характеризующими его свойствами. Представленное изображение на рис. 1 иллюстрирует результат установления взаимосвязи «объект – свойство», сформированной в ходе проведённого исследования. Связь сущности и единичного свойства обозначается одинарной стрелкой, а связь между сущностью и множественным свойством обозначается двойной стрелкой. Также постоянные свойства обозначены непрерывной линией. Если свойство присутствует у всех объектов данного класса, оно обозначается буквой «S».



Рис. 1. Пример изображения получившейся в результате работы связи «объект – свойство»

Помимо связей между объектами и их свойствами, были зафиксированы связи между объектами разных классов (рис. 2). Двойная стрелка говорит о том, что в населённом пункте может проживать несколько объектов населения.

В ходе работы над даталогической моделью выяснилось, что Postgres не предоставляет функциональности для работы с ГИС-данными, поэтому необходимо было установить расширение PostGIS, которое является мощным инструментом для хранения, обработки и анализа пространственных данных, без которого было не обойтись в данном случае. В результате даталогического проектирования было выделено несколько сущностей:

1. Образование. Содержит сведения об учебных заведениях, включая координаты, название, принадлежность к приходу, вид учебного заведения, численность учащихся по полу, количество неправославных учеников, а также подробную информацию об учителях (ФИО, дата рождения, пол, сословие, образование, место получения образования, годовое содержание).

2. Церковь. Детализирует храмы и церкви, включая их названия, приходы, даты просьбы о постройке, закладки, освящения, типы проектов, материальные аспекты строительства и содержания (средства прихожан, частные пожертвования, государственные субсидии), страховые данные (сумму и срок страховки), материалы сооружений, библиографические ссылки, координаты, наличие приделов (наименование, материал, сумма ремонта, характер ремонта), библиотек (количество томов в библиотеке, источники пополнения библиотеки), культовых сооружений, обществ и организаций, корреспондентов «Томских епархиальных ведомостей», а также информацию о наличии попечительства и расходах за год и суммах на конец отчётного периода.

3. Населённый пункт. Хранит данные о населённых пунктах, такие как название, дата основания, причина основания, географические координаты, административная принадлежность (уезд, волость, приход, благочиние), роль населённого пункта

(волостной центр, приходский центр, центр благочиния), тип населённого пункта, количество дворов, площадь земельных участков, удалённость от ключевых объектов инфраструктуры (уездный центр, благочинный, транспортные пути, реки), инфраструктура и окружающая среда.

4. Причт. Содержит информацию о священнослужителях, включающую приход, сан, ФИО, дату рождения, вид и размер содержания, этапы карьеры, награды и назначения.

5. Население. Предоставляет демографические данные населения, включая распределение по полу, этнической принадлежности, вероисповеданиям и направлениям.

6. Религиозные девиации. Собирает сведения о религиозных отклонениях, включая вовлечённые конфессии, виды действий, местоположения событий и библиографические ссылки.

7. Социальные конфликты. Документирует социальные конфликты, их предметы, внутренние и внешние конфликтующие стороны, координаты событий и библиографические ссылки.

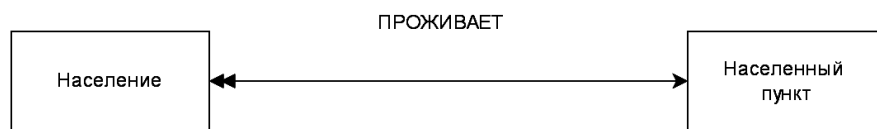


Рис. 2. Пример изображения получившейся в результате работы связи «объект – связь»

В результате проведённых исследований была разработана концептуальная схема базы данных для информационно-аналитической веб-платформы «Православный ландшафт таёжной Сибири», включающая более 40 взаимосвязанных таблиц. Полученная схема демонстрирует сложную систему иерархических связей между разнородными элементами различных тематических областей через центральную таблицу «Населённые пункты» и предоставляет возможность углублённого исследования множества сфер жизни конкретного региона. Такой подход обусловлен тесной корреляцией исторических данных с конкретными географическими локациями, включая городские поселения, сельские общины и прочие типы населённых пунктов.

Отдельно на рис. 3 выделен фрагмент базы данных, описывающий населённый пункт и его отношение к волости, уезду, приходу и благочинию. Дополнительно представлено отношение, обеспечивающее детализацию характеристик населённого пункта в различные периоды его существования. В этом контексте также включены таблицы, отражающие проведённые инспекционные мероприятия в населённых пунктах, включая отдельную таблицу для идентификации лица, возглавляющего проверку, с указанием его должности, причины проведения проверки и типа принятого воздействия по её результатам. В результате все атрибуты отношений являются простыми, а используемые домены содержат только скалярные величины. Каждый ключевой атрибут демонстрирует функциональную зависимость от первичного ключа, исключая возможность декомпозиции. Все неключевые атрибуты

характеризуются отсутствием транзитивной зависимости от первичного ключа. Исходя из вышесказанного, можно заключить, что структура базы данных соответствует требованиям третьей нормальной формы.

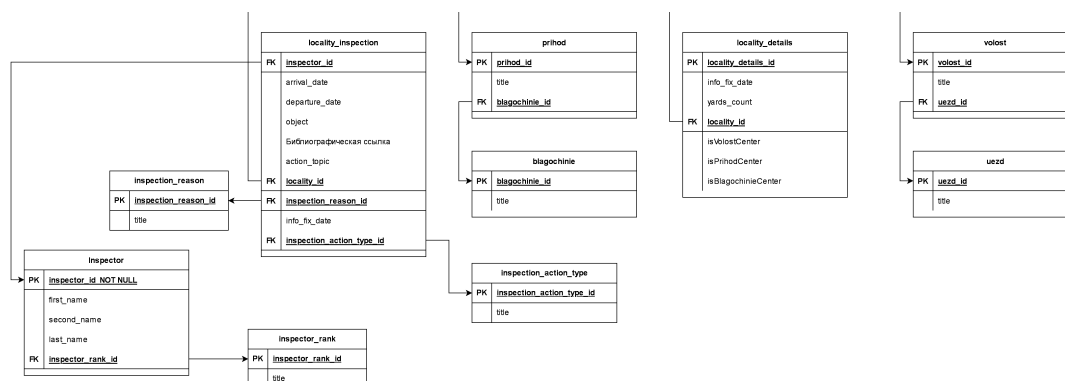


Рис. 3. Фрагмент базы данных информационно-аналитического онлайн-ресурса «Православный ландшафт таёжной Сибири», описывающей населённый пункт

Таким образом, на данном этапе проекта была создана нормализованная база данных, отвечающая требованиям современных информационных хранилищ. Особое внимание было уделено учёту сложных межобъектных связей внутри исторической инфраструктуры. Были идентифицированы основные параметры структурирования данных, что обусловило необходимость модификации классических методик проектирования.

Заключение

В результате проведённых исследований была разработана реляционная база данных, представляющая собой фундаментальную компоненту разрабатываемой веб-платформы, ориентированной на систематизацию, хранение и доступ к историческим данным. Следует подчеркнуть, что процесс проектирования осуществлялся в тесном сотрудничестве с научным коллективом Томского государственного университета, что обеспечило учёт сложных взаимосвязей, характерных для данной предметной области. Наличие большого числа количественных и качественных показателей позволит проводить статистический анализ, а обширные библиографические ссылки делают возможным углублённое исследование исторических источников.

На текущий момент проект находится в фазе активной реализации, включая создание пользовательского графического интерфейса и расширение функционала платформы для проведения аналитических операций. Промежуточные результаты формируют основу для последующего развития проекта и демонстрируют потенциал использования передовых информационных технологий в сохранении и исследовании исторического наследия. Достигнутые результаты служат фундаментом для дальнейшего совершенствования платформы и её применения в историко-культурных исследованиях.

Благодарности

Данная работа была выполнена при поддержке гранта № 23-78-10119, выделенного Российским научным фондом.

Литература

1. Антопольский А.Б. Информационные ресурсы России. Москва: Либерия, 2004. 423 с.
2. Юмашева Ю.Ю. Научное издание исторических документов в электронной среде: проблемы источниковедения и археографии // Письменное наследие и информационные технологии (El'Manuscript-2016) / материалы Междунар. конф. Вильнюс, 2016. С. 334–339.
3. Europeana: исторический информационный ресурс. URL: <https://www.europeana.eu/en> (дата обращения: 20.02.2025).
4. Mapping the Republic of Letters. URL: <http://republicofletters.stanford.edu/> (дата обращения: 20.02.2025).
5. PRO Сибирь: исторический информационный ресурс. URL: <https://prosiberia.tsu.ru/> (дата обращения: 20.02.2025).
6. Илюшечкин В.М. Основы использования и проектирования баз данных: учеб. для вузов. М.: Юрайт, 2024. 213 с.

SYSTEMATIZATION AND PRESERVATION OF HISTORICAL SOURCES IN THE DIGITAL ERA

T.E. Boldovskaya^{1,2}

Ph.D. (Techn.), Associate Professor, e-mail: boldovskaya73@gmail.com

M.I. Remizov^{1,3}

Junior Researcher, e-mail: remizov.maximus.maxim6@gmail.com

A.E. Vetrov^{1,3}

Junior Researcher, e-mail: aleksandrvetrv838@gmail.com

¹Omsk State Technical University, Omsk, Russia

²Dostoevsky Omsk State University, Omsk, Russia

³Tomsk State University, Tomsk, Russia

Abstract. The article is devoted to the problem of preserving historical heritage through the digitalization of written sources. It examines existing electronic repositories and platforms, as well as the process of developing the information-analytical online resource "Orthodox Landscape of Taiga Siberia." The focus is on data unification, the creation of a multi-level storage system, and database design.

Keywords: historical sources, database, data unification, electronic archives, digital humanities, digital heritage, GIS technologies, interfaith relations.

Дата поступления в редакцию: 24.02.2025