

Пермские газеты колчаковского периода: источниковедческий анализ и моделирование информационной системы<sup>1</sup>

Perm newspapers of the «Kolchak period»: source study and the information system modeling

С.И. Корниенко, д.и.н., профессор, профессор кафедры новейшей истории России, руководитель лаборатории исторической и политической информатики ПГНИУ, [kornienko@psu.ru](mailto:kornienko@psu.ru)

Д.А. Гагарина, к.пед.н., доцент кафедры информационных технологий ПГНИУ, [dinara@psu.ru](mailto:dinara@psu.ru)

Н.Н. Масленников, к.т.н., доцент, доцент кафедры информационных технологий ПГНИУ, [nickm@perm.ru](mailto:nickm@perm.ru)

С.В. Пигалева, заведующая сектором книжных памятников Пермского краевого музея, [pokm6@permregion.ru](mailto:pokm6@permregion.ru)

Пермский государственный национальный исследовательский университет (ПГНИУ), Пермь, Россия

Ключевые слова: информационная система, электронный текст, исторические источники, оцифровка.

#### Summary

The article presents the project devoted to preservation, documentation and analysis of Perm newspapers of the «Kolchak period» by means of information technologies. Results of the first stage are given. Conclusions of the source study are drawn. The information model of the collection, its main objects and attributes are described and substantiated.

Актуальность проблемы сохранения, документирования и комплексного изучения коллекций пермских газет колчаковского периода как уникального исторического источника по истории Урала и России и ценного историко-культурного памятника периода Гражданской войны на основе создания информационной системы обусловлена их ограниченной доступностью, малой известностью и недостаточным введением в научный оборот исторических исследований; плохой физической сохранностью, угрозой разрушения бумажного носителя и утраты этих источников; основными

---

<sup>1</sup> Исследование выполнено при финансовой поддержке РФНФ, грант № 11-11-59003-а/У.

тенденциями в развитии информационной среды исторической науки и исторических исследований. Газеты колчаковского периода «Освобождение России», «Свободная Пермь», «Современная Пермь», «Отечество» издавались с 24 декабря 1918 г. по 1 июля 1919 г. – непродолжительный, но значимый период в истории.

Одной из первых задач по созданию информационной системы стал источниковедческий анализ коллекции пермских газет. Он включал в себя выявление сохранности, объема и полноты изданий, их происхождения и характера, структуры, предметно-тематической направленности, лингвистических и полиграфических особенностей.

На основе изучения коллекций газет<sup>1</sup>, библиографических указателей<sup>2</sup>, а также немногочисленных научных трудов по рассматриваемой тематике были составлены библиографические и археографические описания изданий. Составной частью источниковедческого изучения стал формально-логический и предметно-тематический анализ газет. Он показал, что общая структура, названия основных рубрик, тематическая наполненность имеет существенное сходство во всех вышеперечисленных газетах. Структура газет этого периода в целом похожа на Пермские губернские ведомости и другие газеты дореволюционного периода. Это обстоятельство дает возможность использовать опыт ранее созданных на основе дореволюционных газет информационных систем при реализации настоящего проекта.

Изучение полиграфического оформления и состояния газетных коллекций позволило выявить следующее. Сказывался дефицит качественной бумаги, типографской краски, общая спешка в подготовке издания. В силу идеологической направленности истории в советский период коллекции этих газет на долгое время были далеко спрятаны в хранилищах, не ставились на учет, не реставрировались, разрушались. Некачественная газетная бумага,

---

<sup>1</sup> Коллекции сохранились в фондах Пермского краеведческого музея, Государственного архива Пермского края, Пермской государственной краевой универсальной библиотеки им. М. Горького.

<sup>2</sup> Периодика Урала. Вып. 1: Дореволюционные издания: библиограф. указ. / Сост. Г.М. Савиных. Свердловская обл. универс. научн. б-ка. – Свердловск, 1976. – С. 28-30, 122-123; Периодическая печать Урала. Вып. 2: Газеты 1917–1920 гг.: библиограф. указ. / Сост. Г.М. Савиных. Свердловская обл. универс. научн. б-ка. – Свердловск, 1986. – С. 67-81.

угасающий текст – так выглядят сегодня издания, которые отражали повседневность колчаковского периода в Перми, формировали общественное мнение современников. Все это актуализирует проблему перевода данных источников на электронный носитель.

Структурный и предметно-тематический анализ газет позволил сделать вывод о возможности построения общей для всех рассматриваемых изданий информационной модели и создании единой информационной системы. Указанный вывод нашел свое подтверждение и в ходе моделирования каждого из изданий. В качестве основного объекта модели выступает периодическое издание, а атрибутами – название, год, номер, страницы, рубрики, подрубрики, публикации.

Вместе с тем, при информационном моделировании газетной периодики, как показывает опыт создания подобных систем, всегда возникает проблема того, что может быть положено в основу построения модели? Теоретически это может быть и периодическое издание, и его отдельный номер, и страница, и рубрика или подрубрика, и статья. В определенной мере, выбор должен определяться назначением информационной системы и соответствующим подходом к ее созданию – проблемно-ориентированным или источник-ориентированным. При создании достаточно больших и многозадачных систем возможны и комбинации подходов. Однако в данном случае, проектом предусматривается создание источник-ориентированной информационной системы, которая должна дать возможность наиболее полно сохранять и представлять информацию источника. В этом случае наиболее эффективной может стать модель, в основе которой лежит объект «публикация», а остальные свойства издания выступают в качестве его атрибутов. Такая модель позволяет гораздо полнее отразить информацию источника. Более того, именно такая модель наиболее полно обеспечивает информационный поиск в системе на уровне публикаций. В то же время, ее использование влечет избыточность информации, необходимостью постатейного разрезания номеров газет при их переводе в цифровой формат.

В процессе информационного моделирования обсуждались обе модели, в результате был сделан вывод о возможности использования нового

варианта, когда основным объектом, информация о котором будет храниться и отображаться в системе, станет страница номера периодического издания. Остальные элементы станут ее атрибутами. Решение о возможности наиболее полного отражения информации на уровне публикации было перенесено на уровень логической модели, т.е. на уровень организации информации. В ней основным объектом информационной системы является страница номера периодического издания, а номер, рубрики, подрубрики и описание публикации выступают в качестве ее основных атрибутов. Представляется, что такая модель наиболее предпочтительна и с точки зрения технологии оцифровки и распознавания газетных периодических изданий. Она дает возможность размещать основной объект в системе в том виде, в котором он оцифровывается (постранично), без дополнительных весьма затратных операций. Постранично осуществляется в этом случае и распознавание. Для полнотекстового сохранения публикаций и обеспечения работы с ними достаточно копирования, сохранения их текста и его привязки к описанию публикации.

Для реализации модели информационной системы использована СУБД MS SQL на сервере хранилища данных Центра Интернет Пермского университета, с возможностью работы в режиме клиент-сервер и доступом через корпоративную сеть и Интернет.

Таким образом, на основе концептуальной модели предметной области были разработаны технические требования, в соответствии, с которыми создана логическая модель, для ее реализации на физическом уровне была выбрана технология баз данных.

Проделанная работа станет основой для решения дальнейших задач осуществления проекта – непосредственного создания базы данных информационной системы и ввода в нее необходимой информации. Для этого осуществлена оцифровка коллекций газет, постраничное сохранение их каждого номера в электронном формате. В результате решены такие задачи как сохранение ценных исторических источников на новых типах носителя и создание благоприятных возможностей для их дальнейшего использования, изучения традиционными методами. Оцифровка в силу особенностей,

связанных с сохранностью, качеством бумаги и печати, языком изданий, требовала для получения достаточного качества автоматического распознавания использования технологии «распознавание с обучением» и создания пользовательских эталонов.

В ходе решения задач источниковедческого анализа и моделирования информационной системы получили дальнейшую разработку теоретические и прикладные проблемы сохранения и документирования средствами информационных технологий газетной периодики как исторического источника. Разработанные информационные модели источников и системы в целом, методы и технологии моделирования, оцифровки и распознавания обогащают практику создания историко-ориентированных информационных систем и применимы при реализации других подобных проектов.