

ИНФОРМАЦИОННАЯ СИСТЕМА «РОССИЙСКИЕ ПАРЛАМЕНТАРИИ НАЧАЛА XX ВЕКА»

И.К. Кирьянов, С.И. Корниенко

Пермский государственный университет, 614990, Пермь, ул. Букирева, 15

Рассматривается опыт создания информационной системы, ориентированной на решение информационно-поисковых и аналитических задач по истории отечественного парламентаризма начала XX в.

В 2003 г. Российский гуманитарный научный фонд поддержал проект создания информационной системы «Российские парламентарии начала XX века», подготовленный рабочей группой сотрудников историко-политологического факультета Пермского государственного университета. В рамках проекта предусматривалось создание, наполнение, размещение в сети Интернет, с возможностью работы в интерактивном режиме, банка данных о составе и парламентской деятельности депутатского корпуса всех четырех созывов Государственной Думы дореволюционной России, содержащего машиночитаемые данные исторических источников и материалы научного, биографического и библиографического характера и предназначенного для решения информационно-поисковых и аналитических задач.

Информационная система ориентирована на использование метода просопографического исследования – метода «коллективных биографий» или «коллективных портретов». Данный метод создает возможности для многомерного статистического анализа, что позволяет осуществить объективное и достаточно полное исследование, в данном случае – процесса формирования и развития принципиально нового для России начала XX в. слоя политических деятелей – публичных политиков.

В отечественной и зарубежной историографии уже накоплен опыт создания просопографических баз данных. Достаточно упомянуть такие известные проекты, как «Prosopographia regnorum orbis latini» (база биографических сведений из источников III–XII вв.), база данных о членах социал-демократической партии Германии за 1890–1914 гг., база данных для восстановления истории шведских семей на основе приходских книг XIX в. В последние десятилетия идет процесс формирования просопографических баз данных оте-

чественными исследователями. Среди удачных проектов можно упомянуть «Comadam» (просопографическая база данных о высшем командном составе советских вооруженных сил в период Великой Отечественной войны – разработчики Ю.Ю. Юмашева, И.М. Гарскова), «Duma I» (просопографическая база данных о депутатах I Государственной Думы – разработчики Л.И. Бородин, Ю.Г. Григорьева, Н.Б. Селунская), «Workers» (база данных, составленная на основе «личных дел» бакинских нефтепромысловых рабочих «Товарищества нефтяного производства “Братья Нобель”» - разработчики П. Аханчи, И.М. Гарскова). Однако большинство просопографических баз данных носит «персональный» характер и создается отдельными исследователями для решения собственных научных задач. Как правило, персональные базы данных недоступны для других исследователей. Нередко наблюдается ситуация, когда различными исследователями создаются самостоятельные просопографические базы данных по схожим сюжетам и на основании одних и тех же источников. В качестве примера можно привести базы данных о депутатском корпусе Государственной Думы дореволюционной России. Над созданием баз данных по различным созывам отечественного парламента работали в первой половине 1990-х гг. коллективы исследователей исторического факультета МГУ и Пермского университета.

Развитие исследований на основе математико-статистических методов позволяет поставить на повестку дня создание таких машиночитаемых источников, которые были бы многозадачными и многопользовательскими. Путь к этому лежит через создание источникориентированных баз данных, способных функционировать в сети Интернет.

Работа по созданию информационной системы «Российские парламентарии начала

XX века» потребовала осмысления и решения целого ряда источниковедческих, программных, технологических и иных задач, что представляется значимым результатом работы над проектом. Применение более сложных, связанных с использованием компьютерной техники, количественных методов, требует машиночитаемых источников. При этом речь идет не о простом переводе в электронный текстовый формат источников на бумажных носителях. Как показывает опыт, необходимо формирование на основе имеющихся традиционных источников своего рода новых метаисточников, в которых информация организована в соответствии с технологией баз данных. Именно такие машиночитаемые источники позволяют применять сложные статистические методы обработки данных, такие как корреляционный, регрессионный, кластерный, дисперсионный анализ.

В распоряжении исследователей для наполнения базы данных, содержащих социокультурные и политические характеристики российских парламентариев начала XX в., имеются как архивные материалы (массив личных дел депутатов всех созывов хранится в фонде 1278 Российского государственного исторического архива), так и опубликованные источники в виде различных биографических справочников. С учетом того, что практически о каждом депутате можно найти сведения как минимум из трех источников (личное дело, справочник М.М. Боиовича, личный указатель к стенографическим отчетам заседаний Государственной Думы), процедура перепроверки данных значительно облегчена. В результате анализа информативного потенциала источников и технических возможностей их перевода в машиночитаемый формат базовым источником создания информационной системы были выбраны указатели к стенографическим отчетам заседаний Государственной Думы.

Эти документы характеризуются достаточно высоким уровнем структурированности. Основные элементы структуры документов в своей основе едины и сохранены в материалах обо всех четырех думских созывах. В структуре указателей выделяется два главных раздела: «предметный» и «личный». «Предметный» содержит перечень всех рассмотренных в Думе вопросов, сгруппированных по статьям. Объемные статьи делятся на рубрики. Указатели имеют приложения,

включающие достаточно формализованные и унифицированные списки и таблицы: списки членов Государственной Думы по избирательным округам, должностных лиц Государственной Думы, членов Совета Министров, заявлений о запросах (последний снабжен именованным, предметным и алфавитным указателем); таблица законопроектов, внесенных в Государственную Думу; расписание заседаний Государственной Думы.

Для создания машиночитаемого источника, используемого в просопографическом исследовании, особое значение имеет «личный» указатель. Структурирование, формализация и унификация личных данных депутатов в указателях, положенные в основу машиночитаемого источника, обеспечивают возможность решения таких исследовательских задач, как выявление связей между социокультурными и политическими параметрами отдельных депутатов и парламентских фракций, групп в целом.

В то же время справочно-поисковый характер указателей порождает и определенные трудности при создании на их основе машиночитаемого источника. Они связаны с недостаточной в ряде случаев глубиной и полнотой информации. Путь преодоления трудностей лежит через использование материалов других традиционных источников, прежде всего стенографических отчетов заседаний Государственной Думы. Это становится возможным благодаря унифицированному аппарату ссылок на соответствующую информацию в стенографических отчетах и других документах, которым располагают указатели.

Собственно перевод указателей в машиночитаемый формат осуществлялся на основе сканирования источников, представленных на бумажных носителях, с помощью планшетных сканеров и последующего распознавания с помощью стандартной программы сканирования и распознавания Fine Reader Professional 6.0.

При решении этой задачи мы, как и другие исследователи, столкнулись с трудностями, связанными с тем, что источники содержат старорусскую орфографию и грамматику, а также с особенностями шрифтов, которые использовались типографией Государственной канцелярии. В результате сканирования получались достаточно четкие, хорошо читаемые графические образы документов, но

при распознавании, даже в случае применения специально создаваемого алфавита и достаточно длительного распознавания каждого нового пакета-источника, не удалось избежать значительного количества нераспознанных или ошибочно распознанных элементов. Поэтому следовало осуществить правку распознанного изображения, близкую по затратам к ручному набору имеющегося электронного текста. Исходя из этого, было решено отказаться от сплошного распознавания и работать с графическими образами, прибегая к выборочному распознаванию конкретных элементов сканированного источника. Электронные версии документов-источников в качестве архива сохранялись имиджами, что при соответствующем программном обеспечении создает возможность беспрепятственного обращения к информации электронной копии источника. Материалы источников вводились либо прямо в справочники и основные таблицы баз данных, либо предварительно в составляемый машиночитаемый список персоналий, либо в случае необходимости, подвергался фрагментарному распознаванию. На наш взгляд, таким образом удалось оптимизировать затраты на перевод данных источника в машиночитаемый формат.

Важным элементом в работе по созданию информационной системы стало проектирование ее концептуальной и логической моделей. В силу особенностей структуры основного источника – указателей к стенографическим отчетам – эти модели строились на основе выделения двух типов объектов (сущностей) – «Персоны» и «Дума». Генерализующей сущностью ввиду просопографического характера создаваемой системы являлась сущность «Персоны». Сущность «Дума» и ее атрибуты были подчинены задаче наиболее полного и оптимального отображения информации, относящейся к генерализующей сущности. Для построения информационной модели сущности «Персоны» была взята структура информации личного указателя. Для выявления атрибутов сущностей «Персоны» и «Дума» были использованы и другие указатели, списки, таблицы, содержащиеся в основном источнике. Важнейшие свойства этих сущностей были выделены в отдельные атрибуты, в совокупность сущностей второго порядка и их атрибутов. Данные, выступающие в качестве атрибутов различных сущностей, должны были быть унифицированы и

сосредоточены в виде справочников, а также связаны с соответствующими типами объектов соответствующими типами связей – один с одним, один со многими, многие со многими.

Многообразие и сложность концептуальной модели, объектов и атрибутов баз данных, связей между ними, характер данных, цель и задачи информационной системы требовали весьма пристального внимания к выбору языка программирования и СУБД для нее. В качестве языка программирования был избран платформи-независимый язык PL/SQL, а системой управления базами данных – СУБД «Oracle», Oracle Application Server, Oracle Tools. Такой выбор был вызван тем, что на этой базе выполняется ряд научных и образовательных проектов в Пермском государственном университете. Указанный выбор позволил достаточно рационально решить и ряд других задач.

В частности, благодаря созданию и размещению информационной системы на сервере Пермского университета под управлением Oracle Application Server появилась возможность не устанавливать на рабочих станциях специальных клиентских приложений. Создание приложений на PL/SQL и запуск их на Oracle Application Server обеспечили полную кроссплатформенность системы. Вследствие этого появилась возможность уже на этапе создания информационной системы и ввода данных работать через корпоративную сеть университета и Интернет, используя Internet-браузеры. Таким образом, изначально информационная система предполагалась как доступная через Интернет.

Информационная модель системы имеет три основных компонента (раздела): «Персоналии»; «Дума»; «Справочники». Раздел «Персоналии» состоит из основной таблицы, предназначенной для ввода персональных данных о членах Государственной Думы. Раздел «Дума» содержит таблицы, включающие данные о каждом из четырех созывов Думы: список членов Думы соответствующего созыва, даты начала и окончания работы созыва в целом и каждой сессии в отдельности, состав формальных и неформальных думских подразделений (отделы, комиссии, фракции).

Создание информационной системы потребовало решения проблемы стандартизации и унификации данных источников. Особенность ее решения в этом случае состояла в

необходимости учета важной роли системы справочников. Ввод материалов источников в унифицированном виде первоначально осуществлялся в справочники. На соответствующие поля основных таблиц информационной системы данные могли быть введены только из справочников в унифицированном стандартном виде. Одновременно для сохранения уникальной (нерегулярной) информации источников в определенной таблице предусматривается поле «Комментарий», связанное с основным полем и доступное через него. Всего в информационной системе представлено 15 справочников и 18 основных и вспомогательных таблиц.

В ходе развития информационной системы предполагается создание еще трех самостоятельных баз данных, включающих основные источники по каждому депутату, библиографию работ, посвященных думцам, фотографии и иные изображения депутатов, с последующей интеграцией этих таблиц в информационную систему.

По содержанию вся основная информации, которая должна быть отражена в информационной системе, разделится на две части. Одна часть – социокультурные и политические характеристики депутатов Думы (даты жизни, сословная, национальная и конфессиональная принадлежность, характер и уровень образования, профессиональные занятия, имущественное положение, включая владение собственностью, административная деятельность, общественная активность, включая членство в политических партиях), другая – характеристика парламентского поведения (активность при инициировании законопроектов и запросов, а также при их обсуждении, участие в работе думских комиссий, фракционная принадлежность, участие в поименных голосованиях, нарушения думского регламента и др.). Такая логика, а также технология создания информационной системы определили этапы решения поставленной задачи.

В первую очередь проектировались и создавались приложения и интерфейсы для ввода данных в базу. Прежде всего должны были быть введены данные в справочники. После этого на основании заполненных справочников вводились данные в таблицу «PERSON», затем туда шли сведения о социокультурных и политических характеристиках каждого члена Думы, занесенного в указанную

таблицу.

В 2003 г. была завершена отладка всей системы ввода данных и ее проверка. После ввода в базу социокультурных и политических характеристик депутатов была осуществлена проверка этих данных, их редактирование и исправление ошибок. Таким образом, с завершением первого этапа реализации проекта информационная система была частично введена в эксплуатацию, в режиме как корпоративной сети Пермского университета, так и сети Internet. Отдельные элементы ее уже были доступны пользователям.

Главными задачами на втором этапе реализации проекта (2004 г.) стали

- ввод в информационную систему данных о парламентском поведении и думской активности депутатов и в связи с этим проектирование и создание приложений и интерфейсов для ввода и отображения этих данных;

- обеспечение возможности широкого использования информационной системы - как для получения информации о Государственной Думе и ее депутатах, так и для решения задач просопографических и иных научных исследований посредством компьютерной обработки и анализа полученных сведений. В значительной мере решение этой проблемы было связано с проектированием и созданием приложений и пользовательских интерфейсов для соответствующих запросов в базу данных информационной системы;

- обеспечение пользователям возможности доступа и всестороннего использования электронных копий указателей к стенографическим отчетам Государственной Думы дореволюционной России. В связи с этим вставляли вопросы о целесообразности подготовки, проектирования и создания полнотекстовой электронной версии указателей к стенографическим отчетам Государственной Думы, о ее размещении в корпоративной сети Пермского университета и сети Интернет, записи на локальные носители (CD-ROM).

Для решения первой из указанных задач сначала планировалось применить методику и технологию обработки и представления в базе информационной системы материалов, которые уже были использованы при работе с информацией о половозрастных, социокультурных и политических характеристиках депутатов. Однако данные о парламентском поведении и думской активности большинства

депутатов, зафиксированные в личных алфавитных указателях к стенографическим отчетам, не только включают достаточно унифицированные стандартизованные сведения об основных видах депутатского поведения и активности, но и сопровождаются многочисленными ссылками на те или иные тома, разделы и иные фрагменты стенографических отчетов. Причем такие ссылки представляют собой условные кодовые обозначения, требующие расшифровки. Применение уже использованной методики обработки и ввода данных ориентировало на то, чтобы не включать (игнорировать) указанные ссылки (все они представляют собой весьма разнородную информацию). В этом случае характеристики депутатской активности можно было бы получать на основе информационной системы в виде определенных количественных показателей. Например, число случаев голосования «за» или «против», количество выступлений по запросу или законопроекту и т.д. Оценка думского поведения депутатов с учетом качественных характеристик была бы практически невозможна. Для ее осуществления необходимо было либо полностью расшифровать условные обозначения, либо их сохранить в системе данных о думской активности в качестве ключей для последующей работы. Последнее обстоятельство диктовало необходимость применения новой методики ввода в информационную систему материалов о думской активности и их представлении.

С учетом возможностей языка программирования и СУБД был избран такой путь представления данных о депутатской активности, который позволял с максимальной полнотой отобразить информацию первоначального источника. Для этого сканированные фрагменты, содержащие сведения о парламентской активности каждого депутата в каждом созыве Думы и на каждой сессии, подвергались распознаванию и редактированию и были сохранены как отдельный файл HTML. Этот файл связан со специальным полем, отображаемым в интерфейсе для запросов, и позволяет получить данные о депутатской активности в полнотекстовом формате в рамках соответствующего запроса наряду с другими характеристиками. Достаточно хорошо структурированные данные об основных типах думской деятельности депутатов дают возможность пользователям достаточно быстро получить количественные характери-

стики, а условные обозначения открывают возможность выявления качественных, содержательных параметров. С введением данных о парламентском поведении и думской активности депутатов был завершен ввод основного массива данных в базу информационной системы.

Одновременно с решением указанной задачи осуществлено проектирование и создание пользовательских приложений и интерфейсов для различных запросов в базу данных с целью получения информации по персоналиям и проведение научных исследований, их тестирование и отладка.

Результатом этой работы стало создание двух основных пользовательских приложений и интерфейсов. Первое приложение и соответствующий интерфейс рассчитаны на самый широкий круг пользователей и предназначены для предоставления информации о депутатах Государственной Думы по основным социокультурным и политическим параметрам, показателям думской активности, содержащимся в информационной системе. Данный запрос осуществляется на основе выбора соответствующей персоналии.

Другое приложение и соответствующий интерфейс призваны обеспечить информацией, необходимой для просопографических и иных исследований. По сути, это приложение представляет собой конструктор запросов, с помощью которого пользователь может выбрать самые различные критерии, предусмотренные в информационной системе и отражаемые в интерфейсе в случае такого запроса. Данное приложение позволяет осуществлять типологизацию, классификацию депутатов Думы по всем основным показателям, по совместной встречаемости возрастных, географических и иных параметров. Таким образом, с помощью таких запросов могут быть проанализированы состав и характеристики российских парламентариев начала XX в., а также получены материалы для применения других методов обработки и анализа, в том числе математико-статистических.

Как уже отмечалось, немаловажным результатом является проектирование и создание электронной полнотекстовой версии указателей к стенографическим отчетам Государственной Думы I–IV созывов, предназначенной для размещения в Интернете и записи на CD-ROM.

При разработке полнотекстовой версии указателей к стенографическим отчетам были использованы модель и технология публикации «электронная книга». Сканированные изображения основных структурных элементов каждого указателя (оглавления, указатели, таблицы и т.д.) были распознаны, переведены в формат PDF и сохранены отдельными файлами в папке соответствующего указателя. Каждый такой указатель снабжен гипертекстовым оглавлением, которое позволяет переходить к любому структурному элементу внутри указателя. Все указатели имеют общую гипертекстовую страницу, с которой возможен доступ к любому указателю. Создание такого компьютерного продукта существенно расширяет доступ к одному из важнейших источников по истории Государственной Думы в дореволюционной России. Размещение электронной версии указателей к стенографическим отчетам Государственной Думы в Интернете, распространение ее на локальных носителях увеличивает возможности использования электронных ресурсов в научных и образовательных целях. Полнотекстовая версия указателей к стенографическим отчетам, созданная вместе с информационной системой, облегчает исследователям применение методов компьютеризированной обработки и анализа данных, расширяет палитру таких методов. Полнотекстовая электронная версия указателей записана на CD-ROM и размещена в библиотеке электронных публикаций на сервере Перм ГУ.

Основной итог реализации проекта состоит в создании доступной через корпоративную сеть Пермского университета и сеть Интернет информационной системы, представляющей собой источник- и проблемно-ориентированную базу данных. Создана основа для централизованного накопления и многоаспектного использования материалов по истории отечественного парламентаризма

начала XX в. В процессе реализации проекта накоплен опыт решения таких актуальных проблем применения компьютерных технологий в исторических исследованиях, как дигитализация разнообразных источников, развитие технологий создания и использования (в том числе сетевых, источник-ориентированных) баз и банков данных в исторических исследованиях. Проект расширяет доступ к источникам и материалам по истории Государственной Думы, открывает возможность осуществления новых исследований, основанных на использовании количественных методов, таких как

- просопографические исследования с применением корреляционного анализа для изучения социальных портретов думцев, как отдельных созывов, так и депутатского корпуса в целом;

- исследование парламентского поведения с применением кластерного анализа ролевых функций депутатов для выделения типов парламентариев по предпочтительным формам парламентской активности в Думе, а также для изучения становления парламентской элиты в России;

- исследование деятельности парламентских фракций и групп с применением различных методов (корреляционный, кластерный и регрессионный анализ) для изучения механизма принятия решений в предреволюционной России, а также для изучения политических партий и их взаимодействия в политическом процессе начала XX в.

Таким образом, применение количественных методов позволит более полно и объективно исследовать политический процесс в России начала XX в., расстановку политических сил, реальные изменения в политической системе России, происшедшие в связи с реформой государственного строя 1905-1906 гг., функционирование мира публичной политики в России.

INFORMATION SYSTEM “RUSSIAN MEMBERS OF PARLIAMENT IN THE BEGINNING OF THE XX CENTURY”

I.K.Kiryakov, S.I.Kornienko

Perm State University, 614990, Perm, Bukireva, 15

The authors analyze the experience of creation of computer database on composition and activity of pre-revolutionary Russia State Duma deputies designed for working with it in interactive mode.