

СПЕЦИАЛЬНЫЙ ВЫПУСК БЮЛЛЕТЕНЯ

"НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ И ОБРАЗОВАНИИ"

(Тезисы Международного семинара.

Ужгород, 11-14 июня 1992 г.)

Международный симпозиум "High-Tech History Workshop - East and West", состоявшийся в январе 1992 г. в Зальцбурге (Австрия), положил начало активным контактам стран Западной и Восточной Европы в области исторической информатики. Инициатива в развитии этих контактов исходила от Международной Ассоциации "History & Computing" (особую роль здесь играет ее президент, д-р Манфред Таллер). Первым конкретным шагом на этом пути явилось проведение в Ужгороде в июне 1992 г. Международного семинара по использованию новых информационных технологий в исторических исследованиях и образовании. Организаторами семинара являлись Закарпатская ассоциация молодых историков и Исторический факультет МГУ им.М.В.Ломоносова (в лице лаборатории исторической информатики).

В работе семинара приняли участие 40 специалистов из Австрии, Азербайджана, Белоруссии, Великобритании, Венгрии, Нидерландов, Норвегии, Польши, России, Словении, Украины, ФРГ, Чехо-Словакии.

На семинаре обсуждались следующие темы:

- базы данных в исторических исследованиях: достоинства и недостатки работы со стандартным программным обеспечением;
- программное обеспечение, ориентированное на исторические источники;
- экспертные системы и базы знаний в исторических исследованиях;
- новые программы статистического анализа исторических источников;
- компьютерные технологии в анализе нарративных источников;
- компьютерный анализ изображений, гипертекст и мультимедиа;
- компьютеры в обучении историческим дисциплинам;
- проблемы преподавания курса исторической информатики;
- применение компьютерных методов анализа информации в конкретных исторических исследованиях.

В течение трех дней работы семинара были обсуждены около 30 сообщений; многие участники демонстрировали оригинальное программное обеспечение (местные организаторы создали для этого хорошие условия). Для большинства сообщений и дискуссий были характерны конструктивность и заинтересованность. Следует отметить, что программа семинара не предусматривала параллельной работы нескольких секций, так как каждое сообщение представляло интерес практически для всех участников (при этом на каждом заседании был "полный сбор", так что многим не удалось ознакомиться с достопримечательностями Ужгорода). Интенсивности работы семинара способствовало и то, что заседания шли только на английском языке, без потери времени на перевод.

Семинар проходил в обстановке хорошего взаимопонимания между коллегами из разных стран, в дружеской атмосфере не только во время заседаний, но и в редкие свободные минуты. Может быть закономерно, что именно в этой атмосфере возникли несколько плодотворных идей: о проведении осенней школы по исторической информатике в МГУ (при поддержке Международной Ассоциации "History & Computing"), об организации в России ассоциации "История и компьютер". Приятно отметить конструктивность стиля деятельности научного сообщества в области исторической информатики - обе эти идеи уже реализованы: в октябре с.г. успешно прошла первая Московская международная осенняя школа "Историческая информатика: европейская модель", а в ноябре официально зарегистрирована Ассоциация "История и компьютер".

* * *

Данный номер Бюллетеня содержит тезисы сообщений Ужгородского международного семинара, сгруппированные в соответствии с тематикой заседаний. Перевод тезисов на русский язык осуществлен Е.Беловой, Л.Бородкиным, И.Гарсковой, Т.Изместьевой, В.Лазаревым, А.Сафьяновым, Ю.Юмашевой. Техническая работа по подготовке издания проведена с помощью Т.Василевской и С.Ломовой. Редактор выпуска - И.Гарскова.

Редколлегия

ОГЛАВЛЕНИЕ

ИСТОЧНИКО-ОРИЕНТИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Л.Брере. Реляционные базы данных и свободный текст: противоречие в терминах?	5
М.Копчински, Е.Мачак. Специализированный текстовый процессор для ввода данных в SQL/DBMS	13
Э.Костолански. К вопросу о создании интегрированных систем исторической информации (IHIS)	16
В.Леверманн. Данные и знания в исторических базах данных	18
М.Таллер. Что такое "источнико-ориентированная обработка данных" и что такое "историческая информатика"?	20

КОМПЬЮТЕРИЗОВАННЫЙ АНАЛИЗ ТЕКСТОВ

Е.Белова, В.Лазарев. "Система управления историческими текстами". Принципы построения	28
Л.Бородкин, Д.Измestьев, Т. Измestьева. Программа "КОНТЕНТ" для анализа исторических текстов	29
Е.Злобин. Программа статистического анализа текстовых исторических источников	31
Н.Селунская, Л.Бородкин. Базы данных "ДУМА": проблемы комплексирования источников	32

БАЗЫ ДАННЫХ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ

П Аханчи. Источниковедческие проблемы создания баз данных по личным делам нефтепромысловых рабочих г.Баку (на примере фирмы "Товарищество бр. Нобель")	34
Х.Дидерикс, Х.Тьялсма. Компьютер в исследовании профессий	36
С.Журавлев, В.Тяжелникова. Иностранцы в СССР (20-30-е г.). Проблемы формирования базы данных	38
Т.Моисеенко, М.Свищев. Базы данных по новейшей истории сельского хозяйства и крестьянства России (к проблеме вторичного использования)	40
А.Мюллер. Миграция и проблема подвижности населения в средние века (на основе изучения "посетителей" Венского университета)	42
А.Тихонов. Создание базы машиночитаемых данных на основе комплекса разнотипных источников (Бельгийские компании в России конца XIX-начала XX века)	44
К.Шурер. Информационные технологии в исторических исследованиях: взгляд из Великобритании	46
Ю.Юмашева. Источниковедческие проблемы создания просопографических баз данных	48

СТАТИСТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ: НОВЫЕ ПРОГРАММНЫЕ ПРОДУКТЫ

Л.Бородкин, И.Гарскова. FUZZYCLASS: гибкое программное средство построения многомерной типологии объектов социальной природы	51
П.Доорн. Географическое положение, модели взаимодействия и реконструкция исторических поселений и коммуникаций (на примере Этолии, центральная Греция)	54
Я.Олдерволл. WINCENS или новая система для работы с переписями	57
Е.Осокина, Г.Сатаров. Прикладные возможности статистического пакета "CLAMS" (анализ распределительной политики советского правительства	

в первой половине 30-х годов)	58
В.Подгаецкий, Ю.Святец. Программное обеспечение многомерного статистического анализа в задачах изучения социально-экономической истории Украины периода НЭПа	61
Г.Сатаров. Латентные факторы политического размежевания на съездах народных депутатов	62

**КОМПЬЮТЕРЫ В ОБУЧЕНИИ ИСТОРИИ И ПРЕПОДАВАНИИ
КУРСА ИСТОРИЧЕСКОЙ ИНФОРМАТИКИ**

Е.Балыкина. Историческая азбука в компьютерных играх-головоломках	64
Д.Измествев. Система разработки обучающих программ по истории на основе языка SMILE (Story Making Interactive Language)	65
М.Королева, Е.Злобин. Автоматизированная обучающая система по истории России	66
Г.Шпренгнагель. Международный курс "Новые методы в истории" в Зальцбургском университете - развитие дополнительных учебных программ и программ для аспирантов	68

КОМПЬЮТЕРНЫЙ АНАЛИЗ ИЗОБРАЗИТЕЛЬНЫХ ИСТОЧНИКОВ

В.Морозов, А.Пружинин. Средства ввода изображений	71
Ю.Фикфак. Словенская Штирия накануне мартовской революции 1848 года (междисциплинарный проект)	72
Г.Яритц, Б.Шух. Проект базы данных "REAL" по изобразительным источникам	74

ИСТОЧНИКО-ОРИЕНТИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Лен БРЕРЕ
Утрехтский университет, Голландия

РЕЛЯЦИОННЫЕ БАЗЫ ДАННЫХ И СВОБОДНЫЙ ТЕКСТ: ПРОТИВОРЕЧИЕ В ТЕРМИНАХ?

1. Почему системы управления базами данных полезны для историков?

Очевидно, чтобы хранить информацию в памяти компьютера, хотя не менее важен и другой аспект: СУБД полезны, так как требуют высокой степени структурированности данных и поэтому позволяют гораздо легче извлекать информацию.

2. Структура как ключевое понятие хранения и извлечения информации.

2.А. Неэффективное извлечение информации из простого текста можно проиллюстрировать простым примером. Городской архив Гданьска имеет коллекцию писем XVI в. из голландских городов в Городской совет Гданьска по вопросам взаимной торговли и положения эмигрантов. Если мы хотим записать текст письма в память компьютера, это проще и быстрее всего сделать с помощью обычного текстового редактора.

Однако, когда потребуется извлечь из текста какую-либо информацию, возникнет много проблем. Весь текст представляет собой длинную строку символов, а поиск данных сводится к поиску конкретной последовательности символов по всей этой огромной строке. Очевидно, что необходимо просматривать именно весь текст, т.к. он не имеет какой-либо особой структуры, способной помочь в этом поиске.

2.Б. Существуют более эффективные формы извлечения данных, отличающиеся от этой ситуации. Они опираются на введение некоей структуры в хранимую информацию.

(1) Индексация и разметка слов текста. Так называемые полнотекстовые системы индексируют все слова текста, что ускоряет операцию поиска. Например, если нас интересует слово "Амстердам", мы можем указать смысл поиска: порт приписки определенного корабля, отправитель письма, торговый партнер, гражданство и т.д. Тогда в полнотекстовой системе схема поиска может упрощенно выглядеть таким образом:

показать лишь те случаи, где слову "Амстердам" предшествует или за ним следует слово "горожанин", "гражданин" или "капитан" внутри промежутка, например, из 5 слов.

В только что описанной ситуации у нас имеется несколько более структурированный текст, чем в первом примере простого последовательного поиска ключевого слова. Индексирование позволяет рассматривать текст с некоторой структурой: вместо строки символов он становится системой слов (или знаков), которые можно расположить в алфавитном порядке. Так как предполагается, что каждое индексированное слово содержит указатель на то место в тексте, где оно встречается, легко найти и просмотреть нужный контекст. Однако, и этот способ извлечения информации не является во всех отношениях удовлетворительным, так как полный текст не содержит каких-либо четких семантических структур, к которым можно было бы обратиться, а понятие слова для этих целей слишком неопределенно.

(2) Семантическая структура. Ситуация меняется, если мы вводим СУБД. Одной из основных характеристик такого программного обеспечения является его требование к структурированности данных. Структура присутствует на двух уровнях: на низшем уровне информация разбивается на отдельные поля. Семантическим элементом здесь является то, что каждый тип поля может содержать только определенный тип данных. Главное достоинство СУБД, однако, заключается в структурировании данных на более

высоком уровне. Поля группируются в записи, а записи могут связываться друг с другом. Таким образом, база данных может отражать связи объектов реального мира. Возвращаясь к нашему примеру: в одной записи может содержаться информация об определенном письме: дата, отправитель, адрес, номер дела в архиве. В другой записи может содержаться информация о шкипере корабля: имя, фамилия, место рождения, роль в данном письме. Эта вторая запись связана с первой. Третья запись требуется для описания его корабля, а эта запись, в свою очередь, связана с записью о шкипере. Теперь наш запрос может стать гораздо более конкретным. Можно точно указать тип данных, которые надо включать в этот запрос:

показать все записи о шкиперах, имеющие слово "Амстердам" в поле "место рождения", и для них выдать всю информацию о письме и грузе корабля.

Кроме точности, при поиске мы располагаем и дополнительной эффективностью СУБД: используется не только индексирование информации, но и точное "знание" области для поиска, т.к. просматривается содержимое лишь определенных полей текста.

2.В. Обратная сторона эффективности. Вводя структуру, мы, однако, создаем не очень гибкий инструмент, который может существенно помешать свободе обработки данных, особенно учитывая богатство исторических источников. Историк обычно имеет дело с данными, не структурированными по форме, и стремится к вводу в базу данных свободного текста. Это происходит при обработке исторической информации так часто, что проблема заслуживает систематического подхода, а не быстрого, но частного программного решения. Перед тем, как перейти к практическим деталям, сформулируем эту проблему в более общем плане.

3. Создание базы данных: модель реальности.

Благодаря структурированию данных, наша база способна отражать определенные стороны внешнего мира. Более строго, мы организуем данные согласно некоторой модели этого мира. В этом контексте модель понимается как упрощенный образ реальности, созданный под определенным углом зрения. Я буду называть модель данных моделью реальности, чтобы отличить ее от других видов моделей, которые будут описаны ниже. Планирование базы данных, таким образом, называется моделированием данных. Однако, мы говорим также о модели реляционной базы данных которая, строго говоря, также является моделью данных, однако на ином уровне. Последнюю можно называть "конструктивной" или формальной моделью (см. ниже). Поэтому, чтобы не возникло путаницы, будут использоваться оба термина: модель реальности и формальная модель.

Модель данных переводится на язык базы данных, создаваемой в рамках системы управления базами (СУБД). Ее можно рассматривать как основу специальной базы данных; как некую схему, показывающую конструкцию создаваемой базы данных и рациональные пути поиска данных в базе. Разумеется, историк, в противоположность своим коллегам из других сфер деятельности, не может легко верифицировать свою модель, т.к. историческая реальность локализована в прошлом.

Моделирование исторических данных - это деятельность, требующая профессионального мастерства историка, подобно переводу средневековых календарных дат в современную систему, интерпретации водяных знаков, знанию денежных единиц разных эпох или умению читать старые рукописные документы.

4. Правила структурирования: формальная модель.

Структурирование (или моделирование) данных можно сравнить с игрой, имеющей строгие правила. Правила игры всегда имеют две стороны: с одной стороны они поддерживают и направляют нашу деятельность, с другой - ограничивают свободу действий. Системы управления базами данных сильно различаются между собой по набору таких правил. Мы будем называть такой набор формальной моделью или

моделью отношений. Каждый пакет СУБД строится на такой модели, и для того, чтобы правильно построить базу данных, эту модель надо четко представлять. Другими словами, формальная модель определяет тот способ, который позволяет представить некоторую модель реальности в виде базы данных. Приятно, если правила игры просты и их можно освоить за короткое время. Возможно, это одна из главных причин того, что реляционная модель приобрела такую популярность в последнее десятилетие.

Реляционная база данных состоит из отдельных таблиц, каждая из которых есть набор записей, которые можно упорядочить разными способами. По сравнению с иерархической или сетевой реляционная модель является плоской, т.е. все таблицы (все записи) находятся на одном уровне. Однако, можно легко превратить иерархическую (или сетевую) модель реальности в реляционную базу данных. Для этого реляционная модель отношений имеет механизм т.н. ключевых полей: таблицы можно связывать друг с другом по совпадающим значениям этих полей. Вернемся к нашему примеру с письмом, шкипером и грузом его корабля: эту модель реальности можно отразить в трех связанных таблицах. Запись о шкипере должна содержать ключевые значения записи о письме, где упоминается этот шкипер. Аналогично, запись о корабле связывается с записью о шкипере. Одной таблицы было бы недостаточно, т.к. в одном письме могут упоминаться несколько шкиперов, а один шкипер может быть связан более, чем с одним кораблем.

5. Ограничения реляционной модели.

Реляционная модель базы данных хорошо известна своей прочной математической базой и дружественным пользовательским интерфейсом. Однако на практике историки могут встретиться с определенными трудностями. Как уже отмечалось выше, не всегда бывает легко сформировать общее представление о материале источника и разработать такую модель базы данных, которая во всем устроила бы исследователя. Большинство часто встречающихся трудностей можно свести к различным видам нестандартности данных, вполне обычным для исторических источников. Они отражают изменения во времени и пространстве административных и социо-культурных систем, а также уникальность исторических событий. Имена могут иметь непредсказуемую длину, даты могут относиться к разным календарным системам, налоги быть частично денежными, а частично натуральными, возраст иногда дается лишь в виде интервала, а некоторые данные требуют длинных комментариев.

В таких ситуациях кажется заманчивым "скорректировать" данные источника при вводе, однако, т.к. историки занимаются своими поисками в прошлом, любая модификация источника может привести к потере информации. Следовательно, источник-ориентированная обработка данных требует, чтобы машиночитаемая копия источника была как можно более близкой к оригиналу. А это неизбежно приводит к заключению, что применение пакетов реляционных баз данных не может обойтись без проблем при попытке следовать стандартам реляционной модели.

Есть разные пути преодоления разрыва между богатством исторической реальности и жесткой конструкцией реляционной модели: либо полностью отказаться от последней и обратиться к более ориентированным на историю системам баз данных, подобным KLEIO Манфреда Таллера, либо вводить специальные средства в программное обеспечение коммерческих пакетов реляционных СУБД. Под специальными средствами имеются в виду программы, обеспечивающие решение различных проблем в обобщенной форме, т.е. они не ограничены одним-единственным источником или проблемой, но могут использоваться в самых разных ситуациях, а дружественный интерфейс скрывает от неопытных пользователей технические детали. В следующем разделе это утверждение будет рассмотрено подробнее на примере превращения свободного текста в записи реляционной базы данных.

6. Преобразование свободного текста в реляционную базу данных.

Важным аспектом компьютерной технологии в истории является обработка документов, содержащих фактические данные. Многие источники официального происхождения содержат текстуальные описания таких данных. Можно привести много примеров: письма, уставы, нотариальные акты, документы о правах на владение, завещания, описи имущества и т.д. При работе с ними историка в основном интересует фактическая информация. Документ, содержащий эту информацию, можно свести к такой модели данных (в смысле модели реальности), которая подходит для реляционной базы данных. Как правило, такие данные можно правильно понять только в исходном контексте, значит, имеет смысл ввести в базу данных и соответствующий отрывок текста, а еще лучше - весь документ. Контекст - сложное понятие, и во многих случаях трудно выделить его достаточно четко. Насколько далеко надо идти по тексту, зависит от цели создания базы данных и природы источника. Для простой книги регистрации рождений хватит обычной реляционной таблицы. Однако, если имеется много записей на полях книги, а база данных предназначена для использования с разными целями, можно предусмотреть и более сложную организацию информации.

В принципе для таких задач реляционные СУБД подходят не слишком хорошо. Из-за фиксированных размеров полей и принципа индексации они требуют данных ограниченной длины. Основные операции реляционной алгебры, такие, как селекция, проекция, объединение, предполагают именно такую структуру. Пока для решения этих проблем в коммерческих пакетах предлагаются лишь длинные символьные поля (или поля MEMO) для комментариев, что никак нельзя назвать универсальным решением, т.к.: обычно с такими полями не разрешены никакие операции, кроме просмотра; эти текстовые данные нельзя включать в поиск вместе с фактическими данными; приходится искусственно разбивать исходный документ на части, а это разрушает концепцию контекста; наконец, документ не только разбивается на фрагменты, но эти фрагменты ограничены одной таблицей (файлом базы данных).

Таким образом, нам нужна некая система, помещающая весь документ (или его фрагменты) в память машины в виде общего файла, доступного одинаково из каждого файла базы. Кроме того, этот текст должен допускать какую-либо возможность поиска вместе с фактическими данными.

* * *

В рамках исследовательского проекта SOCRATES в Утрехтском университете в данное время разрабатывается набор средств обработки данных, содержащихся в документах. В качестве реляционной СУБД мы используем dBASE IV, расширив ее командный язык с помощью библиотеки программ, содержащей около 100 процедур и функций. Когда библиотека подключена, любую из этих процедур можно вызвать через меню или из любой другой программы.

7. Этап 1. Редактор документа.

7.А. Обзор. Наши специальные средства включают редактор документа, позволяющий одновременно работать со свободным текстом и структурированными данными в оболочке реляционной СУБД. Во-первых, этот редактор можно использовать как обычный текстовый редактор: данные обоих типов легко вводить, модифицировать и уничтожать, имеются и другие возможности, присущие текстовым редакторам, например, импорт и экспорт ASCII-файлов.

Однако, при более близком знакомстве можно заметить некоторые возможности, не свойственные стандартному текстовому редактору. Файл с документом является не обычным ASCII-файлом, а файлом базы данных. Каждая запись состоит из одного текстового поля такой длины, которая вмещает одну строку текста, и нескольких полей для целей управления. В одном из них содержится метка, указывающая способ использования упомянутого длинного текстового поля. Это длинное поле имеет две функции: или в нем записана строка обычного текста или оно разделено на виртуальные поля.

7.Б. Виртуальные поля. Эти поля создаются автоматически, включением в длинное символьное поле некоторого шаблона. Тип шаблона соответствует структуре записей файла базы данных. Каждый элемент шаблона используется для ввода данных определенного типа, например, имени, фамилии, места рождения и т.д. и в этом отношении подобен обычному полю базы данных, однако фактически это некая симуляция полей, почему они и называются виртуальными. Вообще говоря, этот подход не отличается от разметки блоков текста с автоматическим добавлением некоего кода или от выделения в тексте каких-то данных с добавлением явного кода. Достоинством нашего метода является то, что данные, записанные в виртуальных полях, не обязательно должны непосредственно встречаться в тексте в явном виде. Впоследствии эти поля автоматически копируются в файлы реальной базы данных; на этом этапе редактор документа работает как буфер данных. После любого параграфа текущего текста можно вставить сколько угодно шаблонов (см. рис. 1 и 2).

TAG		LINE			
PERSON		Cameron		John	
				19201205	
					шаблон
					виртуальные поля

Рис. 1. Пример шаблона для имени, фамилии и даты.

Чтобы пометить контекст, пользователь должен добавить некоторые метки, каждая из которых является отдельной строкой. Строка должна быть вставлена в какое-то место в столбце Jump_label, идущем по всему тексту документа (имя Jump_label станет яснее при описании операции извлечения текста). Jump_label можно сравнить с точкой для ориентира. Как уже сказано, контекст в нашем понятии не есть нечто статичное, следовательно для разметки контекста вместо жестких границ мы используем такие "маяки".

Таким образом, документ хранится в файле, имеющем длинное текстовое поле LINE. В этом поле может стоять строка текста или структурированные данные. Поле TAG сообщает, что именно содержится в поле LINE (см. рис. 2).

TAG	LINE			
	с человеком, по договоренности			текущий текст: поле
	с его хозяином из Данцига...			TAG пусто
SOURSE	15260221	письмо	Staes Goverts	данные: поле TAG
PERSON	Staes		Goverta	сообщает тип этих
PERSON	Herman		Doctoer	данных

Рис. 2. Основной экран редактора документа

После ввода документа и разметки всех фактических данных с помощью шаблонов можно поместить курсор на некоторый шаблон, нажать клавишу - и данные из всех шаблонов этого типа будут скопированы в соответствующий файл базы данных. Соответствующая метка перехода Jump_label копируется в соответствующей файл базы вместе с данными.

8. Этап 2. Контекстный поиск.

Для доступа к данным библиотека программ SOCRATES имеет пользовательский интерфейс, несколько более совершенный, чем стандартный интерфейс dBASE, предлагаемый режимами EDIT или BROWSE. На экране имеются 3 области: средняя часть используется для ввода данных или поиска и вывода информации очередной записи. Нижняя часть содержит меню с функциями редактирования. В верхней части располагается основное рабочее меню. Это меню содержит возможности вызова

нескольких интерфейсов, включая не только стандартные EDIT и BROWSE, но также и возможности доступа к данным, существующие лишь в системе SOCRATES. Среди них - JUMP.

Выбрав область данных, можно "листать" базу. Предположим, что мы находимся в файле базы данных, который перед этим был автоматически заполнен в процессе копирования данных из виртуальных полей в файле документа. Для какой-то из записей этого файла потребовался исходный контекст. Для этого надо вернуться к рабочему меню и выбрать пункт JUMP. Тогда специальная программа считает "индекс маяка": имя файла и метку, которая была записана в текущую запись вместе с данными. Пользователь "перепрыгивает", если можно так выразиться, назад к контексту документа. На экран накладывается окно с началом соответствующего абзаца, который можно просмотреть вверх и вниз. По контрасту с полями типа MEMO, этот фрагмент текста не ограничен.

9. Применение: расширение формальной модели.

Суммируем наши заключения в нескольких утверждениях:

1) Реляционная модель структуры лишь в незначительной степени может работать с неоднородной информацией источника. В принципе структура записи фиксирована, а таблицы связаны, и мы должны учитывать эти связи. Наша свобода ограничивается возможностью пропускать связанные записи, но если данные уже введены, то переопределять отношения нельзя.

2) Обычно модель реальности стремится к простоте и ориентирована на стандартизованную информацию. Для справедливости скажем, что это хорошо не только для реляционной СУБД, но и привлекательно для пользователя, способного оценить такую простоту.

3) Однако, историческая реальность часто более сложна и требует корректной работы с нестандартизованными данными. Предлагаемое решение проблемы касается создания набора специальных программных средств, выходящих далеко за пределы стандартно понимаемой реляционной модели, но не нарушающих обычных операций реляционной алгебры. Набор программных средств не должен усложнять объединение баз, требовать низкого уровня нормализации или каким-либо иным образом делать базу менее реляционной. Этот набор должен лишь добавлять новые средства, не отказываясь ни от одной из стандартных возможностей.

4). Введение свободного текста в реляционную систему использовано в качестве примера того, как записи одной и той же таблицы могут быть связаны нерегулярным образом с записями других таблиц. JUMP - это действительно универсальный способ, который необязательно ограничен проблемой контекстного поиска. Адрес перехода может быть сохранен в любой записи, и JUMP откроет файл, указанный в этом адресе, там, где находится метка. Каждая последующая запись может, разумеется, указывать на другой файл, а это, очевидно, совсем не то, что стандартная связь между таблицами, основанная на ключевых полях.

5) Концепция файла документа демонстрирует, как можно хранить текст в реляционных таблицах без полей типа MEMO и без ограничений на длины полей (подробное рассмотрение этих структур не является, однако, целью данной статьи).

Михал КОПЧИНСКИ
Варшавский Университет,
Ежи МАЧАК

Технологический Университет, Варшава, Польша

СПЕЦИАЛИЗИРОВАННЫЙ ТЕКСТОВЫЙ ПРОЦЕССОР ДЛЯ
ВВОДА ДАННЫХ В SQL/DBMS

Проект, который мы хотим представить, базируется на регистрах подушной подати 1662, 1673 и 1676 годов. Это - единственный источник достаточной эпохи, на основе которого можно анализировать структуру хозяйства или собственности, а также изучать социальную структуру населения по всей территории Польши. Его достоинства были отмечены еще в начале 60-х годов, но возможности фундаментальной обработки и анализа этого огромного по объему источника открылись только с появлением компьютеров.

В начале 1990 года данные этого источника были введены в компьютер. Мы надеемся, что база данных записей подушных податей послужит началом масштабного электронного архива, который в конечном счете будет создан совместно с другими польскими историками.

Текст источника вводится в компьютер в исходном виде. Затем он маркируется с помощью индексов, подобных меткам SGML (Structured Generalized Markup Language - Язык Структурированных Обобщенных Меток). Индексы подразделяются на группы, связанные с определенными особенностями индексируемого текста, такими, например, как описание налогоплательщика, членов его семьи или других проживающих совместно с ним. Каждая группа имеет свои собственные индексы, относящиеся к детальным характеристикам описываемых персон. Так же, как и индексы, группы могут быть вложенными, но не должны перекрываться.

Каждый элемент регистра рассматривается отдельно и индексируется словом <регистр>. Программа присваивает каждому элементу ввода уникальный номер. Комментарии, сделанные оператором, заключаются в скобки {" " "}. Если в этом есть необходимость, две группы индексов (например, группы индексов, относящиеся к описанию дворянина и его собственного слуги) могут быть связаны.

Маркировка текста является сложной работой, особенно если обрабатываются тексты, содержащие, подобно регистрам подушных податей, много информации. Чтобы упростить эту работу, мы разработали программу, которая содержит функции обработки текста и создания базы данных посредством преобразования индексируемого текста источника. Помимо элементарных функций текстового процессора, таких, как чтение и сохранение файла, удаление, копирование, поиск и замена, программа имеет набор функций, обеспечивающих индексирование текста. Среда пакета WINDOWS 3.0 позволяет с помощью "мыши" легко маркировать индексируемые фрагменты текста. Библиотеку индексов можно ввести в оперативную память компьютера с дискового файла (ASCII-файл), легко модифицируя его, если необходимо, как на время текущего сеанса работы, так и с целью последующего использования. Это возможно, поскольку множество индексов не является закрытым и его можно увеличить (добавлением индексов) или уменьшить (уничтожением индексов).

Индексирование текста проводилось не с клавиатуры, а с помощью "мыши", при этом соответствующие группы и типы индексов отбирались из меню. Программа автоматически вставляет начальный и конечный индексы в текст источника, однако сначала подходящие индексы необходимо ввести в библиотеку. Другие функции могут проверить корректность индексирования отобранного фрагмента текста, записи или файла. Посредством специальной команды индексы или комментарии пользователя можно удалить из текста.

На основе анализа индексируемого текста можно также генерировать SQL-таблицы. Язык SQL (Structured Query Language - Язык Структурированных Запросов) представляет собой набор хорошо организованных вопросов, понятных компьютеру. Все данные хранятся в системе данных SQL в форме таблиц. Каждая таблица состоит из столбцов (по вертикали) и строк (по горизонтали). По аналогии с традиционными базами данных строки SQL-таблицы можно рассматривать как записи, а столбцы в каждой строке как поля записей.

Завершенная база данных - это набор таблиц. Таблицы могут быть как независимыми, так и связанными друг с другом принудительно или общими полями (например, номером записи). Принудительная связь определяется отношением между внешним ключом в зависимой и первичным ключом в родительской таблицах.

Осуществляемая нами генерация SQL-таблиц в форме ASCII-файлов с описанием структуры позволяет импортировать их в базы данных, поддерживающие возможности языка SQL, т.е. dBase IV на PC или VM SQL/DS на больших компьютерах.

Программа оперирует исключительно с ASCII-файлами, которые могут быть подготовлены с помощью специального встроенного текстового процессора или любого текстового процессора, имеющего экспорт в ASCII-файл. Она не создает никаких дополнительных внешних структур данных. Результатом работы программы является индексированный текст источника в виде ASCII-файла и наборы текстов, содержащих таблицы баз данных. Эти таблицы, если они содержат столбцы чисел, могут быть введены в статистический пакет, например в SPSS.

Авторам программы хотелось бы в дальнейшем привести индексированные текстовые файлы к стандарту SGML (ISO 8879).

Приложение

Фрагменты текста источника и индексированного текста

ASK I/71, k.128

Rostoka

Generosus {Pietkiewicz} subdapifer bielski, a officio 12 zl,
a persona sua 3, generosa consors eius 1.15; a tenuta 45,
generosa Zamoyska 1.15, synowie 2 zl. 3, corka 1 zl. 1.15,
famuli nobili 2 zl. 4; pers. pleb. curialibus 13 zl. 13.

<REGISTER 1>

<SIGNATURE> ASK I/71 </SIGNATURE>, k. <PAGE> 128 </PAGE>

<PLACE_NAME> Rostoka </PLACE_NAME>

<OWNER>

<TITLE> Generosus </TITLE> <FAMILY_NAME> {Pietkiewicz}

</FAMILY_NAME> <OFFICE> subdapifer bielski </OFFICE> a

officio <CHARGE_FOR_THE_OFFICE> 12

</CHARGE_FOR_THE_OFFICE>

zl., a persona sua <CHARGE> 3 </CHARGE> <SEX> {m} </SEX>

<FAMILY>

<TITLE> generosa </TITLE> <KINSHIP> consors {zona}

</KINSHIP> eius <CHARGE> 1.15 </CHARGE> <SEX> {k}

</SEX>

</FAMILY>

</OWNER>

<OTHER_DUES> a tenuta </OTHER_DUES> <CHARGE> 45 </CHARGE>

<RESIDENTS>

<TITLE> generosa </TITLE> <FAMILY_NAME> Zamoyska

</FAMILY_NAME> <CHARGE> 1.15 </CHARGE> <SEX> {k} </SEX>

<FAMILY>

<KINSHIP> synowie <KINSHIP> <NUMBER> 2 </NUMBER>

zl. <CHARGE> 3 </CHARGE> <SEX> {m} </SEX>

</FAMILY>

<FAMILY>

<KINSHIP> corka </KINSHIP> <NUMBER> 1 </NUMBER> zl.

<CHARGE> 3 </CHARGE> <SEX> {k} </SEX>
</FAMILY>
</RESIDENTS>
<SERVANTS>
<APPELLATION> famuli </APPELLATION> <ORIGIN> nobili {s}
</ORIGIN> <NUMBER> 2 </NUMBER> zl. <CHARGE> 4 </CHARGE>
</SERVANTS>
<SERVANTS>
<APPELLATION> pers. pleb. curialibus </APPELLATION>
<ORIGIN> {p} </ORIGIN> <NUMBER> 13 </NUMBER> zl.
<CHARGE> 13 </CHARGE>
</SERVANTS>
</REGISTER 1>

Эдуард КОСТОЛАНСКИ
Братиславский университет, Словакия

К ВОПРОСУ О СОЗДАНИИ ИНТЕГРИРОВАННЫХ СИСТЕМ ИСТОРИЧЕСКОЙ ИНФОРМАЦИИ (IHIS)

1. Интегрированные Системы Исторической Информации (Integrated Historical Information Systems - IHIS). Все более глубокое проникновение компьютеров и коммуникационной технологии (чаще используется термин "информационная технология" - ИТ) будет способствовать появлению IHIS, основным признаком которых является то, что они будут иметь дело не с документами, а с информацией и знаниями. Следует отметить, что переход от документов к знаниям не является простой и линейной процедурой. Можно представить себе возможности и преимущества IHIS, но варианты развития этой системы менее очевидны. Следовательно, необходимо уточнить содержание IHIS и пути развития этих систем. Основными проблемами при создании IHIS являются накопление информации и знаний во всем их объеме и сложности, а также связь с пользователями.

Разработка IHIS должна способствовать развитию нескольких научных направлений:

а) Конструирование IHIS - это вызов для специалистов по информатике и базам знаний, проверка способности соответствующих областей науки решать сложные задачи моделирования различных типов информации и обработки этой информации в реальном времени.

б) IHIS проверяет систематику описания и представления информации и знаний из данной исторической области в связи с их компьютерной обработкой. В этом отношении развитие могло бы состоять в установлении стандартов, унификации терминологии, элиминирования избыточности и т.п.

в) Прямой доступ к огромным базам знаний, выход с их помощью на различные уровни детальности, установление различных зависимостей должно стимулировать творческий потенциал пользователя, т.е. способствовать активизации конкретно-исторических исследований.

2. Характеристика IHIS и проект создания. IHIS принадлежит к группе таких информационных разработок, для дальнейшего развития и совершенствования которых полезно и выгодно не задавать заранее жестких ограничений. Разработка IHIS могла бы развиваться по спирали. Вот две типичные черты спирального развития информационного продукта:

а) Отдельные части в цикле создания информационных продуктов (спецификация, концептуальная схема, компьютеризованная модель, содержащая операции и

инструментальные средства) разрабатываются не последовательно, шаг за шагом, но итеративным путем. Первые витки спирали концентрируются на спецификации, последние - на инструментальных средствах.

б) Вторая характерная черта спирального развития предполагает обязательное участие пользователя на всех стадиях создания информационного продукта. Пользователь нужен в самом начале, поскольку он, безусловно, лучше знает возможные приложения создаваемого продукта. Участие пользователя на протяжении всего процесса разработки гарантирует, что не возникнут искажения или неверное понимание первоначальных намерений.

3. Коммуникация с IHIS. Обратим внимание на две важные операции. Первая называется "силуэтинг" (silhouetting), вторая - фокусирование (sharpening). Силуэтинг должен давать сообщение о том, существует ли вообще объект, который нас интересует, и краткую информацию о нем. Фокусирование представляет более детальную информацию об этом объекте.

4. Вместо заключения. Мы надеемся, что пришло время совместного проектирования IHIS.

Вольфганг ЛЕВЕРМАНН
Институт истории им.Макса Планка, Геттинген, Германия

ДАННЫЕ И ЗНАНИЯ В ИСТОРИЧЕСКИХ БАЗАХ ДАННЫХ

Хотя название этой статьи наводит на мысль о противоречии между данными и знаниями, тема ее безусловно противоположна, а именно - их синтез. Тема эта, конечно, не новая, хотя обычно упомянутые понятия принадлежат совершенно разным "мирам" исследований в computer science: в то время как первое относится к базам данных (БД) и системам управления базами данных (СУБД), второе тесно связано с искусственным интеллектом (ИИ) и более популярной областью его практического применения - экспертными системами и базами знаний.

Тем не менее, в последнее время интеграция методов ИИ с технологией БД стала очень популярной темой исследований среди специалистов по БД. С другой стороны, потребность в практических решениях на основе баз данных все больше и больше осознается специалистами по базам знаний. В результате постепенно исчезают различия между БД и БЗ, и обе технологии участвуют в создании многочисленных практических приложений.

Основная тенденция такого развития, видимо, имеет большое значение для историков, использующих компьютер как инструмент исследования. При работе с историческими источниками для правильного их понимания и интерпретации обычно приходится привлекать большой объем исторических данных. Следовательно, компьютерная поддержка историка в его изысканиях должна заключаться не только в использовании больших массивов данных, но и в предоставлении дополнительных средств, которые помогли бы ему в понимании источника. Безусловно, значение методов, имеющих целью интеграцию знаний с управлением данными, особенно посредством технологии БД, будет возрастать.

Вообще, существует довольно большое число решений проблемы сочетания баз данных и экспертных систем, и наиболее интересные из них уже реализованы. Обратимся к некоторым способам, позволяющим объединить внешние исторические знания (historical background knowledge) с управлением ими в рамках такого подхода к обработке данных, который учитывает специфику исторического источника (так называемая "источнико-ориентированная обработка данных" (source-oriented data processing)). В частности, здесь описывается подход, реализованный в системе KLEIO -

программном обеспечении, ориентированном на работу с историческими базами данных (см. статью М.Таллера в этом же сборнике).

Частично отражая традицию научного изучения и издания источников, которая способствовала превращению истории в XIX веке в академическую дисциплину, а частично благодаря практическому опыту работы с пакетами программ, перегружающих процесс ввода исходных данных необходимостью их кодирования, классификации и стандартизации, источник-ориентированный подход нацелен на сохранение максимума информации источника на этапе ввода данных в компьютер.

Хотя для применения такого подхода есть все основания, сохранение оригинальной информации влечет за собой ряд проблем, связанных с ее автоматической обработкой. Максимальное сохранение информации оригинала и по возможности минимальное кодирование ведут к тому, что в большинстве случаев исходные данные представлены не в той форме, которая требуется для их дальнейшего использования и обработки (например, при поиске, сортировке, связывании записей и применении статистических методов).

Представление информации в оригинальном виде даже при простейших операциях поиска данных может вызвать проблемы. Например, если компьютер должен найти все документы, созданные в определенное время, то мы бы предпочли иметь возможность строить запрос к этому периоду, не принимая во внимание различные способы записи даты в историческом источнике. То есть, если мы не хотим модифицировать наши данные при вводе, то необходима дополнительная информация, позволяющая интерпретировать исходную и приводить ее к "стандартизованной" форме или - в общем случае - к такому способу представления, который наиболее соответствует дальнейшей обработке. В принципе, это может быть сделано двумя способами: дублированием информации путем ввода "вручную" добавочных полей или атрибутов в базу данных, которая содержит информацию в форме, необходимой для последующей обработки. Второй способ направлен на программную поддержку преобразования данных: при этом функции посредника между оригинальными данными и всеми типами программ, которые используются для их анализа, выполняет программное обеспечение специального типа, содержащее правила конвертирования данных.

В этой статье обсуждаются все "за" и "против" для обоих подходов, однако предпочтение отдается второму. В этом случае "оригинальные" данные и внешние исторические знания существуют как отдельные компоненты базы данных, которые администрируются независимо друг от друга и связываются каждый раз, когда требуется дополнительная информация. Хотя могут возникнуть разногласия относительно того, можно ли рассматривать правила, необходимые для "простых" преобразований дат или денежных единиц, как "знания" в смысле искусственного интеллекта (всегда связанные с некоторой долей предположительности), несомненно, мы входим в эту сферу по меньшей мере тогда, когда программное обеспечение указанного типа расширяется и становится более гибким, нацеленным на более адекватный способ представления исторических знаний в компьютере и связывание их с источниками.

Шагом в этом направлении является так называемая контекстно-чувствительная обработка данных (*context-sensitive handling of data*). Это наиболее тщательно продуманный путь извлечения "смысла" из элементов данных до передачи их другим программам для дальнейшей обработки.

"Смысл" элементов текста формируется динамически посредством интерпретации их в контексте, т.е. путем учета значимой информации из контекста и интеграции экспертного исторического знания. Как следствие, два элемента данных, являющихся полностью идентичными по их внешним характеристикам, могут различно трактоваться системой в зависимости от контекста, в котором они встретились. Предполагается, что в предельных случаях контекстно-чувствительная система будет обрабатывать элемент данных, содержащий строку "белое", так, как будто он содержит термин "черное", если

это следует из контекста. Хотя этот подход отражает довольно тривиальное понимание того, каким важным фактором для интерпретации элементов данных является контекст, он весьма важен при использовании баз данных в гуманитарных науках, поскольку стандартные (коммерческие) СУБД не обеспечивают адекватного решения проблемы.

Манфред ТАЛЛЕР
Институт истории им.Макса Планка, Геттинген, Германия

ЧТО ТАКОЕ "ИСТОЧНИКО-ОРИЕНТИРОВАННАЯ ОБРАБОТКА ДАННЫХ" И ЧТО ТАКОЕ "ИСТОРИЧЕСКАЯ ИНФОРМАТИКА"?

1. Что такое источник-ориентированная обработка данных?

Чтобы плодотворно использовать компьютер в историческом исследовании, необходимы два совершенно отдельных шага:

- извлечение и формализация информации источника (кодирование или "маркирование");
- анализ данных либо с применением инструментария более формализованных дисциплин (подобного статистическим методам или лингвистическим моделям), либо с преобразованием оригинала в форму, более пригодную для интерпретации (подобную, например, тематическим картам).

Компьютер может быть использован на обоих этапах; но мы будем различать методо-ориентированный и источник-ориентированный способы его применения.

Мы говорим о методо-ориентированном подходе, если преобразование источника в набор кодов, представляющих собой строго определенные содержательные категории, полагается только на интеллект исследователя. При этом подходе два сходных термина будут представлены одним кодом, который и вводится в компьютер.

При источник-ориентированном подходе, мы, напротив, всегда вводим в компьютер текст оригинала, но подразумевается, что историк должен в этом случае располагать рабочей средой, в которой может прежде всего выяснить контекст, чтобы принять решение, следует ли при анализе слить близкие по смыслу термины в единую категорию или нет. Такой подход имеет некоторые преимущества: опечатки в словах менее опасны, чем в кодах, проблемы надежности также упрощаются. Однако, гораздо более важно, что данные, введенные в исходном виде, могут использоваться с разными целями. Например, ввод географического названия в качестве значения переменной "место рождения" позволяет нам: преобразовать это географическое название в условный код расстояния; преобразовать это название в пару координат, которая затем может быть использована в вычислениях, основанных на метрических расстояниях; использовать его для построения алфавитного списка мест рождения; использовать транскрипции названий для анализа изменения употребления разных языков с течением времени; строить тематические карты.

Рассмотрим понятие источник-ориентированной обработки данных несколько более формально. Современные аналитические пакеты, подобные SAS, имеют развитые возможности преобразования строковых переменных в числовые коды, но очевидно, однако, что инструкции, указывающие на необходимость интерпретировать "Четверг после Пасхи 1564" как "25 мая 1564 года", гораздо более сложны, чем простая перекодировка. Учитывая это, определим источник-ориентированную обработку данных как попытку смоделировать на компьютере все множество содержащейся в источнике информации, чтобы оперировать источниками для решения возможно более широкого спектра задач.

Рассмотрим реализацию этого подхода в рамках конкретной системы KLEIO, которая разработана автором в тесной кооперации с несколькими исследовательскими группами.

2. Источнико-ориентированная система управления базами данных KLEIO.

Чтобы смоделировать всю совокупность информации, содержащейся в историческом источнике, и управлять ею в вычислительной системе, нужно преодолеть некоторые практические трудности. Наиболее тривиальные из них - это:

- Поля, переменные (т.е. базовые единицы информации) сильно варьируются по длине и могут в одном случае содержать одно слово, а в другом - текст из многих строк.
- Набор базовых единиц информации может не совпадать для различных объектов в источнике.
- Исторические источники (например, завещания) могут быть весьма сложны по структуре, полное воспроизведение которой может потребовать тысяч - а то и десятков тысяч - полей различных типов.
- В то время, как в статистических системах подразумевается, что переменная может принимать одно и только одно значение, исторические источники часто используют много терминов для описания одного понятия, например, профессии.
- В то время, как базы данных (и статистические системы) предполагают, что поля или переменные содержат информацию, связанную с точно определенным понятием, исторические источники часто содержат информацию, которая лишь весьма отдаленно связана с определенным понятием. Следовательно, нужен механизм связывания целых блоков информации, с любым полем в базе данных.

KLEIO использует понятие элемента или элементарной информации. Элементы часто содержатся в совершенно неупорядоченных структурах, которые иногда могут быть описаны, как иерархические. Задача усложняется тем, что для различных типов исследования могут быть полезны различные способы структурирования исторического источника. Исходя из высказанных соображений, KLEIO базируется на оригинальной модели данных, тесно связанной с понятием семантической сети. Эта модель может быть схематично описана следующими основными характеристиками:

- Каждый из объектов в базе данных (называемый группой) может включать произвольное подмножество всех элементов базы данных.
- Каждая группа может быть связана с произвольным числом других групп как в той же базе данных, так и в других.
- Если группы двух различных типов делят элемент с одним и тем же именем, система рассматривает обе как объект одного типа и автоматически осуществляет их "неявное соединение".
- Поскольку многие исторические источники имеют неявную иерархическую структуру, программные средства поддерживают такие отношения и переводят их в более простые.

Между традиционными структурами данных и теми, что описаны выше, существует огромное различие. Реляционная база, например, создается путем заполнения системы таблиц, что имеет существенные недостатки в случае источнико-ориентированной обработки данных. Эти недостатки связаны с фундаментальным отличием исторического исследования от любого другого рода деятельности, подразумевающего обработку информации. Историки не распоряжаются информацией в том смысле, как это делают наши коллеги в естественных науках: если наши источники не содержат необходимой нам информации, мы не можем и не должны пытаться переписывать их. Поэтому модель данных, которую мы описали выше, использует совершенно иную метафору: мы обозначаем этот акт, как "перевод" или "редактирование" источника для читателя не-человека, подобно переводу исторического источника для читателя, который не знает языка, и нам приходится давать дополнительные разъяснения. В результате, описание структуры данных в KLEIO по существу никогда не завершается, т.к. установлены

правила, позволяющие вам вводить дополнительные элементы непосредственно во время ввода данных без явного изменения описания базы данных.

Различие между нашим подходом, реализующим источниковую ориентацию, и методо-ориентированным решением становится еще более очевидным, когда мы смотрим на содержимое полей. Как интерпретировать даты, заданные в нестандартных формах, сравнивать имена, близкие по написанию, или переводить денежные единицы? Конечно, эти проблемы порождаются не компьютером: все эти вопросы точно также возникли бы, пользуясь мы просто карандашом и бумагой. Для решения многих вопросов нам понадобились бы всевозможные записные книжки, справочники, картотеки и т.п., чтобы фиксировать все наши предположения о значении отдельных фрагментов нашего источника.

Именно такую модель KLEIO и пытается реализовать. Это значит:

- Для всех упомянутых проблем, т.е. для хронологии, денежных систем, систем мер и весов, сравнения имен, программа содержит базовые алгоритмы оперирования с информацией такого рода. Например, KLEIO знает, что выражение вида <день_недели><направление_отсчета><праздник><год> есть допустимый способ указания календарной даты.
- Специфические правила, которые должны быть употреблены для данного источника, пользователю приходится указывать самому с использованием соответствующих элементов командного языка системы. В нашем примере мы должны были бы указать, на каком языке будет даваться <день_недели> и когда отмечался день данного святого в регионе, к которому относится наш источник.

Ситуацию, к которой мы приходим, можно представить следующим образом:

Пользователь
Программное обеспечение
База знаний
Транскрипция источника

В KLEIO работа с историческим источником осуществляется путем по возможности буквального переписывания различных фрагментов его, сопоставленных отдельным элементам базы данных. Все знания о значении переписанных фрагментов образуют отдельный слой системы, который специально посвящен управлению знаниями. Любой запрос пользователя, любая команда доступа к записанным данным источника интерпретируется в соответствии со знаниями об источнике, хранимыми в компьютере.

3. Расширение модели.

До сих пор мы описывали основу парадигмы источник-ориентированной обработки данных в истории (в противоположность методо-ориентированной). В действительности в рамках источник-ориентированной методики существуют два различных подхода, к которым в ближайшем будущем, похоже, присоединится третий, вбирающий в себя последние достижения компьютерной технологии.

Первый из них обычно называют структурированным. Он подразумевает, что исследователь присваивает отдельным полям базы данных различные фрагменты данных из источника. Возьмем, например, текст:

В день святого Михаила на сороковой год правления Елизаветы, нашей славной Королевы, Джонни Тернер, что держит кузницу за церковью, уплатил семь гиней своему брату Фредерику, которые тому задолжал, согласно последней воле своего отца, и все его имущество отныне свободно от всяких обязательств.

В KLEIO мы бы ввели его в виде:

платеж\$дата=день Св.Михаила 1598#день Св. Михаила 40 Елиз.И/
сумма=7г/остаток_долга=нет/причина=завещание
плательщик\$фамилия=Тернер/имя=Джонни/занятие=кузнец#
держит кузницу за церковь
получатель\$имя=Фредерик/отношение=брат

На уровне абстракции можно сказать, что в этом случае сохранена семантика текста, синтаксис, однако, утерян: различные управляющие символы сообщают программе, какая часть текста соответствует имени, какая уплаченной сумме и т.д.. При втором подходе те же данные были бы введены в следующем виде:

(платеж (дата В день (праздник святого Михаила на (год сороковой год) год правления (правитель Елизаветы правитель) дата),нашей славной Королевы, (плательщик (имя Джонни имя) (фамилия Тернер фамилия), что (состояние держит кузницу за церковь состояние) плательщик), уплатил (сумма семь гиней сумма) своему (получатель (отношение брату отношение) (имя Фредерику имя) получатель), которые тому задолжал, согласно (причина последней воле своего отца причина), и все его имущество отныне (остаток_долга свободно от всяких обязательств остаток_долга) платеж).

На уровне абстракции мы снова можем сказать, что семантика источника сохранена. Синтаксис, однако, в этом случае также сохранен и даже выделен различными вставками, заменяющими специальные символы предыдущего примера.

Различие между двумя примерами на первый взгляд может показаться незначительным. На практике, однако, оно имеет далеко идущие последствия. Многие исследовательские проекты ставят целью совместить публикацию источника с подготовкой того же текста к компьютерной обработке, что оказывается довольно естественным в парадигме источник-ориентированной обработки данных. И все же при использовании подхода, который мы назвали структурированным, эти две цели не могут быть совмещены полностью: реальный текст источника не может быть восстановлен из базы данных, при использовании же второго подхода (который можно назвать предварительным редактированием) это очень легко сделать.

Для нашей попытки систематического описания источник-ориентированной парадигмы важны, однако, не столько эти практические соображения, сколько общее понятие о привязке структур баз данных вышеописанного типа к полной транскрипции самого исходного текста. Оно так важно, поскольку позволяет легко интегрировать в обобщенную источник-ориентированную парадигму развивающиеся в настоящее время методы обработки изображений. Мы употребляем термин обработка изображений в очень широком смысле, включая сюда четыре основных типа приложений:

- Непосредственный доступ (immediate retrieval) к изображениям, позволяющий в качестве результата операции поиска по базе данных, содержащей описания изображений, получать на дисплее само это изображение.
- Усиление изображений, объединяющее методы улучшения качества поврежденных фрагментов изображений.
- Распознавание образов, в ближайшем будущем, по-видимому, более пригодное для сравнения и классификации, чем для собственно распознавания форм.
- Методы, обычно объединяемые под названием "гипертекст", которые мы, однако, предпочитаем называть связыванием изображений, позволяющие устанавливать связи между частями структурного описания изображения и фрагментами самого изображения.

Большое значение, которое мы придаем этим методам, связано с обработкой специфических изображений: факсимильных репродукций рукописных источников. Такие репродукции могли бы в корне изменить понятие публикации. Сегодня под "изданием" мы понимаем воспроизведение набора источников, чаще рукописных, но обычно в форме транскрипции, являющейся результатом осмысленной интерпретации. Эта транскрипция (вместе с необязательным факсимильным воспроизведением) сопровождается различного рода справочным аппаратом, дающим - в идеальном случае - возможность понять все предположения, сделанные публикатором при составлении транскрибированного текста из рукописных.

При использовании обозначенных выше методов понятие публикации может стать более динамическим. Вслед за традиционным этапом сбора как можно большего количества качественных репродукций источника, встает задача перевода источника в форму оцифрованного изображения. После этого воспроизведение становится практически чисто механическим. В то же время такая форма хранения позволяет применять методы, названные нами усилением изображений и распознаванием образов, и во многих случаях сканированное изображение обрабатывать гораздо легче, чем исходную бумажную копию. Интеллектуальную работу публикатора - определение того, как следует читать и понимать текст, - при этом легко проделать при помощи методов непосредственного доступа к изображениям и связывания изображений, которые позволяют постоянно переключаться с интерпретации текста публикатором к воспроизведению оригинала в рукописной форме. Мы приходим к следующей модификации схемы, реализуемой KLEIO:

Пользователь
Программное обеспечение
База знаний
Транскрипция источника
Оцифрованное изображение источника
(digital image of source)

Пополним наше общее определение KLEIO: В KLEIO работа с историческим источником осуществляется путем по возможности буквального переписывания различных фрагментов его, сопоставленных отдельным элементам базы данных. Все знания о значении переписанных фрагментов образуют отдельный слой системы, который специально посвящен управлению знаниями. Любой запрос пользователя, любая команда доступа к записанным данным источника интерпретируется в соответствии со знаниями об источнике, хранимыми в компьютере. Репродукция транскрибированного источника хранится в базе данных; фрагменты его могут воспроизводиться на подходящем устройстве вывода в результате доступа к транскрипции любым из описанных выше способов. После того, как репродукция части источника выбрана и выдана на экран, она может быть модифицирована подсистемами, включающими весь спектр средств обработки изображений.

4. Что такое историческая информатика?

Мы начали нашу статью с двух вопросов; на первый мы если и не ответили, то по крайней мере прояснили его. Второй в принципе требует гораздо более полного обсуждения: автор уже не раз пытался сделать вклад в такое обсуждение. Здесь автор надеется дать ключ к возможному ответу, рассматривая его как примечание к определениям, которые мы использовали до сих пор. Мы исходим из предположения об общей предпочтительности описанного нами подхода к обработке данных. Чтобы воплотить его, требуются безусловно значительные технические знания. Часть из них - вроде умения пользоваться текстовым редактором для переписывания исторического источника - довольно тривиальны; другие - например, понимание того, каким образом

должны устанавливаться связи между фрагментами изображения рукописного текста и частями описывающей его структуры, - весьма сложны. Большинство лежит посередине - например, поиск путей определения структуры, отражающей содержание сложного документа.

Смесь задач, за разрешение которых должны взяться историки будущего, заставляет нас рассмотреть роль человека, который помогает историкам применять те решения, которые доступны в рамках существующего программного обеспечения, к их задачам. Это роль человека, который применяет методы формального анализа к источниковому материалу, как формализуя содержание источников таким образом, чтобы они могли стать объектом механической интерпретации, так и выявляя структуры, специфические для исторических источников, делая возможным описание общих решений проблем их обработки на современных компьютерах. И наконец, это роль человека, который создает работающие программные решения для тех проблемных областей, где не помогают коммерческие, инженерные и естественнонаучные подходы.

Знания, требуемые для выполнения этой роли, мы называем предметом исторической информатики.

КОМПЬЮТЕРИЗОВАННЫЙ АНАЛИЗ ТЕКСТОВ

Евгения БЕЛОВА, Валерий ЛАЗАРЕВ

Московский государственный университет им.Ломоносова, Россия

"СИСТЕМА УПРАВЛЕНИЯ ИСТОРИЧЕСКИМИ ТЕКСТАМИ". ПРИНЦИПЫ ПОСТРОЕНИЯ

Перевод полных текстов исторических источников в машиночитаемый вид в настоящее время признается насущной потребностью для различных направлений исторического исследования. Представляется необходимым (и возможным - если принять во внимание современный уровень компьютерной технологии) обсудить возможные подходы к построению полнотекстовых исторических баз данных, которые обеспечивали бы эффективное извлечение и анализ всей информации, содержащейся в тексте источника. Мы считаем, что предписывание определенным элементам текста раз и навсегда заданных "ролей" (маркирование текста) ограничивает исследователя, т.к. требует предварительной интерпретации содержания. Напротив, необходима полная семантическая обработка текста, в результате которой устанавливается соответствие между значимыми элементами текста (как правило - словами) и семантическими категориями. Категории должны объединяться в иерархические структуры и структуры типа семантических полей, образующие в целом семантическую сеть. Элементарная информация могла бы извлекаться при этом "по вертикали" (при движении от общих категорий к более конкретным) и "по горизонтали" (к смежным категориям).

Исследователю должна быть дана возможность расширить или перестроить сеть, дабы привести ее в соответствие с привычной ему системой понятий, принятыми предположениями и целями анализа. Вообще говоря, система категорий должна рассматриваться как полностью определяемая пользователем, однако на практике она несомненно содержала бы некоторое стандартизованное ядро. Ради возможности дальнейшего статистического анализа должны быть предусмотрены средства редукции текста к традиционному представлению "переменная * объект = значение".

Определенные элементы текста могут ссылаться на конкретные объекты (факты) реальности, описываемой источником (лица, географические пункты и т.п.). Должны существовать процедуры, которые распознавали бы текстовые элементы как такого рода ссылки. Объекты одного типа должны объединяться в "индексы", причем совокупность последних образовывала бы своего рода систему координат, к которой был бы привязан каждый фрагмент текста. Индексы могли бы быть реализованы прежде всего в форме списков (например, персоналий). Кроме того можно говорить и о непрерывных индексах - шкалах (например, календарных дат, если мы рассматриваем последние, как объекты особого рода - точки на временной оси).

Поддержка системы индексов вела бы к автоматической "гипертекстизации" банка текстов, позволяющей, например, рассматривать несколько текстовых фрагментов, как различные свидетельства об одном и том же событии, поскольку они ссылаются на один и тот же - решающий - набор значений индексов (дата, место, действующие лица).

Так как предлагаемые процедуры опираются на понятие "элемент текста", должны существовать правила описания типов таких элементов и методы их распознавания.

Предполагается, что разрабатываемая в настоящее время в Лаборатории исторической информатики МГУ им.Ломоносова программная система будет обладать некоторыми из описанных черт. Система включит в себя специализированный лингвистический процессор (ориентированный на тексты декларативного характера), анализатор структуры текста, подсистему операций над семантической сетью и подсистему запросов, включающую, в частности, функцию сведения текстов к таблицам данных выбранного формата.

Леонид БОРОДКИН, Дмитрий ИЗМЕСТЬЕВ, Тамара ИЗМЕСТЬЕВА
Московский государственный университет им.Ломоносова, Россия

ПРОГРАММА "КОНТЕНТ" ДЛЯ АНАЛИЗА ИСТОРИЧЕСКИХ ТЕКСТОВ

Формализованный анализ содержания текстов и документов представляет большой интерес для историков. Для решения этой проблемы обычно используется контент-анализ.

Условно процедуру контент-анализа можно разделить на три этапа. Сначала происходит выделение важнейших характеристик текста (индикаторов) и объединение их в содержательные с точки зрения исследователя понятия (категории). Второй этап включает обработку полученных на первом этапе данных с помощью математических методов. Заключительная стадия анализа - интерпретация результатов.

Авторы исходили из представлений о том, что выявление первичных смысловых единиц текста и их укрупнение - прерогатива исследователя. Это творческая, с трудом и не до конца формализуемая деятельность. Разработанная нами (на языке программирования СИ) программа "Контент" (версия 1.0) ориентирована на реализацию основных аналитических процедур контент-анализа.

Какие возможности предоставляет "Контент" исследователю? Программа позволяет вводить индикаторы на естественном языке и кодирует их. Текстуально совпадающие выражения кодируются одинаково. Конечно, это не решает проблему установления смыслового тождества индикаторов. В дальнейшем предполагается включить в программу процедуру выявления семантического равенства различных языковых форм, положив в основу методы, разработанные в структурной лингвистике. В первой версии программы, где эта проблема еще не решена, пользователю предоставляется возможность "насилъственно" присвоить одинаковые коды в случаях совпадения смыслового содержания различных индикаторов. После введения индикаторов пользователь программы может сформировать категории.

Показатели, позволяющие анализировать содержание текстов и документов, рассчитываются как для индикаторов, так и для категорий. Это частоты встречаемости (как на множестве документов, так и на совокупности индикаторов или категорий), коэффициенты связи Юла (Q), сопряженности (Ф), коэффициенты Ласвелла.

Программа проста в использовании. Все операции (ввод данных, редактирование, получение нужных расчетов, вывод результатов) осуществляются с помощью меню. Продуманный интерфейс позволяет сосредоточиться на содержательной работе. Имеется справка, дающая представление об используемых методах, и контекстно-зависимая помощь.

Для удобства пользователя результаты расчетов представлены компактно (для обозначения столбцов и строк в матрицах коэффициентов используются коды), но не в ущерб наглядности (указав курсором на любой коэффициент, можно увидеть, для каких признаков он рассчитан).

Первая версия программы "Контент" освобождает исследователя от известной рутинной работы. Ведется работа по созданию следующей, усовершенствованной версии, более полно отвечающей нуждам исследователей, анализирующих содержание исторических текстов и документов.

Евгений ЗЛОБИН

Гуманитарная Академия Вооруженных Сил, Москва, Россия

ПРОГРАММА СТАТИСТИЧЕСКОГО АНАЛИЗА ТЕКСТОВЫХ ИСТОРИЧЕСКИХ ИСТОЧНИКОВ

Одним из способов внедрения новых компьютерных технологий в исторические исследования является создание специализированного программного обеспечения для персональных ЭВМ (ПЭВМ), ориентированного на особенности исторических источников.

Несколько лет назад автором был разработан пакет прикладных программ для изучения текстовых исторических источников. С его помощью был произведен анализ "Записок" декабриста И.И.Горбачевского. Результатом статистического исследования был вывод о том, что автором анализируемых "Записок" был не И.И.Горбачевский, а другой, неизвестный автор.

В пакете была реализована методика Н.А.Морозова, основанная на подсчете частоты встречаемости служебных слов в предложении. Кроме того, проводился морфологический разбор предложений в полуавтоматическом режиме путем кодирования соответствующих грамматических единиц. После этого просчитывалась частота их взаимной встречаемости в контексте простого предложения. По итогам подсчетов строился граф связей - своеобразная модель текста. Графы двух текстов сравнивались между собой. Пакет был написан на языке Бейсик для микроЭВМ "Искра 226". Помимо чисто статистических функций он позволял также вводить текст в память ЭВМ в виде одномерного массива, каждый элемент которого являлся одним словом. Сейчас этот пакет переработан для IBM-совместимых персональных компьютеров. Созданная программа статистического анализа текстовых исторических источников является одним из элементов автоматизированного рабочего места историка. Эта программа анализирует любой русский текст, представленный в формате ASCII-файла, то есть, может работать с такими текстовыми редакторами, как NB, Лексикон, Multi-Edit и др.

Программа осуществляет поиск и статистический подсчет частоты встречаемости предлогов, союзов и прочих служебных слов в тексте. Общее число таких слов не ограничено, но в данной версии программы составляет около 250. Возможен также поиск слов, целенаправленно задаваемых исследователем. Число таких слов сейчас равно 42, но может быть увеличено.

Один из режимов работы программы - сравнительный анализ двух текстов. После расчета статистических параметров обоих текстов проверяется гипотеза о неслучайности различия их статистических характеристик с использованием критерия Стьюдента. Конечный результат выдается на принтер, дисплей или в ASCII-файл. Работа программы по обработке текста сопровождается музыкой.

Дальнейшее совершенствование программы ведется с целью автоматизации процедуры морфологического разбора предложения. Источниками для исследований являются массивы последних писем и статей Б.В.Савинкова, подготовленные для обработки А.В.Быстровым, а также т.н. "Платформа Рютина" - крупнейший антисталинский документ 30-х годов. Задачи исследования - проверка гипотез о возможном авторстве этих документов.

Наталья СЕЛУНСКАЯ, Леонид БОРОДКИН

Московский государственный университет им. Ломоносова, Россия

БАЗЫ ДАННЫХ "ДУМА":

ПРОБЛЕМЫ КОМПЛЕКСИРОВАНИЯ ИСТОЧНИКОВ

Историки ряда стран Европы в 80-х годах приступили к созданию баз данных, являющихся важной частью исследований по истории парламентаризма. С конца 80-х годов такая работа ведется и на кафедре источниковедения исторического факультета МГУ (исследовательский проект "Дума"). Одна из главных целей проекта - сбор и анализ данных о "коллективной биографии" представителей законодательной власти России

начала XX века - уникального периода "конституционализма", связанного с функционированием четырех Государственных дум.

В центре внимания нашей работы на данном этапе находятся вопросы методики создания базы данных и апробации направлений анализа и обобщения биографических сведений о депутатах I Государственной думы (1906 г.), отложившихся в различных комплексах источников в процессе подготовки и функционирования первого органа общероссийского представительства. Это нормативные документы, определяющие порядок выборов и представительства в Думе; статистические сведения в общероссийском масштабе о численности и структуре населения на губернском и уездном уровне; сведения о нормах представительства различных социальных слоев; делопроизводственные материалы (в том числе протоколы и стенографические отчеты заседаний Государственной думы). В центре внимания находилась публикация "Государственная дума первого призыва. Портреты, краткие биографии и характеристики депутатов" (М., 1906) и комплекс родственных изданий.

Одной из главных проблем разработки компьютерной базы данных "Дума" (реализованной на основе СУБД dBASE III+) является адекватная формализация сведений, почерпнутых из описанного выше комплекса источников и дающих разнообразную по форме и различную по степени информативности, детальности информацию о 478 депутатах I Государственной думы. Решение этой проблемы потребовало систематизации собранных сведений, создания на их основе своеобразного формуляра, насчитывающего более 60 информационных полей. Наибольшее внимание на данном этапе потребовала проблема формализации жизненных путей депутатов Думы. Наряду с динамическими по характеру данными о карьере депутатов структура базы включает большое число сведений демографического и социального характера (возраст, национальность, вероисповедание, сословно-социальные характеристики, данные об образовании); об данных об общественно-политической активности (партийной принадлежности, участии в выборных органах, политической ориентации на разных этапах деятельности депутатов) и др.

При создании структуры базы данных особое внимание уделялось тому, чтобы в максимальной степени учесть все сведения о депутатах, содержащиеся в различных источниках. Так, для характеристики партийно-политических позиций каждого депутата отведены 10 информационных полей, которые отражают партийность депутата и его общественно-политические убеждения, зафиксированные в "Таблице депутатов", в издании "Члены Государственной Думы", в "Портретах", в издании "Наши Депутаты" М.Бойовича и в словаре Брокгауза и Эфрона.

Участие депутатов в общественно-политических изданиях отражено в пяти информационных полях. В тех случаях, когда эта форма активности представлена обширным материалом, вводится соответствующий комментарий (в поле "МЕМО", хотя здесь следует отметить, что информационный поиск по таким полям в реляционных СУБД невозможен, разрешена только процедура DISPLAY для просмотра).

Перспективы работ по проекту Дума состоят в создании последовательности баз данных по материалам этого представительного органа всех четырех созывов за период 1906-1917 гг. Кроме создания коллективного портрета российских парламентариев предполагается разработка целого ряда конкретно-исторических и методологических проблем политической истории России начала XX века, т.е. эпохи конституционализма, а именно: исследование процесса формирования многопартийной системы через процесс функционирования фракций в Думе, перехода к демократии на основе анализа системы выборов, истории электората, феномена политической культуры законодательной элиты России и др., а также разработка методик анализа текстов дебатов, зафиксированных в стенографических материалах, построение методик изучения карьеры депутатов Думы, процесса профессионализации законодательной элиты России.

Решение этих задач потребует использования новых информационных технологий (например, программного обеспечения "полнотекстовых" баз данных), а также "продвинутых" методов репрезентации и анализа текстов (когнитивный подход и др.)

БАЗЫ ДАННЫХ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ

Парвин АХАНЧИ

Институт истории Азербайджанской академии наук, Баку, Азербайджан

ИСТОЧНИКОВЕДЧЕСКИЕ ПРОБЛЕМЫ СОЗДАНИЯ БАЗ ДАННЫХ ПО ЛИЧНЫМ ДЕЛАМ НЕФТЕПРОМЫСЛОВЫХ РАБОЧИХ г.БАКУ (на примере фирмы "Товарищество бр. Нобель")

Квантификация и компьютеризация процесса исторического познания расширяет не только арсенал методов исследования, но и границы вводимых в научный оборот типов источников, в первую очередь массовых. Обращение такого нового направления исторических исследований, как историческая информатика, к созданию баз и банков данных позволяет на качественно новом уровне моделировать источник. Трансформация данных источника в базу данных (БД), являясь частью "нового источниковедения", улучшает качество хранения данных, ускоряет и облегчает их поиск, избавляет историка от части рутинной работы. Особенность исторических баз данных - в максимально полном и максимально близком к оригиналу сохранении информации исторического источника с возможностью последующего многократного и многоаспектного ее анализа для решения различных исторических проблем.

С конца 80-х гг. накоплен интересный опыт по структуризации разных типов источников и переводу их в базы данных. Становится все более популярным обращение исследователей к стандартному программному обеспечению. Имеющиеся в активе историков БД можно классифицировать в зависимости от исходного источникового материала, как текстовые (по нарративным источникам), статистические и формулярного характера (в частности, просопографические). Особо следует остановиться на БД, построенных на источниках формулярного характера, какими являются, например, документы первичного учета (личные карточки, личные дела, анкеты, биографические справочники и т.д.) Данная группа БД отличается наиболее стандартными принципами формализации данных источника даже при разных целях создания, задачах поиска или методах обработки. Кроме того, БД, сформированные на источниках формулярного характера, различаются объектом отражения. Это, во-первых, могут быть персональные анкеты крупных политических или государственных деятелей (членов Государственной думы, делегатов партийных съездов, высшего генералитета Вооруженных сил), т.е. личные данные известных персоналий, а во-вторых, личные карточки рабочих, служащих, и др., так сказать, массовых слоев населения. БД, созданные на основе документов первого уровня, носят характер в первую очередь информационно-поисковых. БД, построенные по документам первичного массового учета, в основном ориентированы на статистическое обобщение (структурный анализ, выявление обобщенных факторов и т.п.).

Личные дела рабочих "Товарищества бр. Нобель" относятся к документам именно такого типа и представляют собой уникальный источник для изучения социально-экономического положения рабочих в нефтяной промышленности Баку в конце XIX-начале XX вв. (хронологический "срез" данных включает 1879 - 1921 гг.). Источник представляет собой комплекс документов первичного учета, отличающийся помимо большого разнообразия и высокой репрезентативности первичной информации, также и временной протяженностью, т.е. динамичностью многих показателей (сведения о многократных поступлениях и увольнениях каждого рабочего, изменения в размерах зарплаты, квартирного и продуктового довольствия, пособия и штрафы, перемещения по службе, несчастные случаи и т.п.) в отличие от многих источников также формулярного характера, но содержащих лишь единовременные сведения. Например, в источнике нашли отражение различные социально-политические события ("военное пособие" - 1914

г., "пособие на дороговизну" - 1915 г., удержания "в пользу раненых", "на содержание лазарета" и т.д.). Наиболее интересна сплошная динамика данных о зарплате, содержащихся в т.н. "Ведомостях", причем важно, что эти сведения даны как в денежном, так и в натуральном выражении.

Процесс формализации источника и процесс его анализа с использованием технологии баз данных, ставит перед исследователем ряд специальных проблем:

- полного или выборочного воспроизведения содержания источника в машиночитаемой форме;
- кодировки исходных данных, особенно с учетом возможности объединения нескольких источников в один, по одним и тем же персоналиям;
- создания одного или нескольких файлов данных, особенно для источников со сложной структурой;
- корректировки возможных ошибок, содержащихся в источнике;
- агрегирования информации с учетом выбора наиболее адекватных количественных методов обработки.

С учетом специфики источника и дальнейшего анализа с помощью статистических пакетов по личным делам рабочих "Товарищества" создана реляционная база данных, состоящая из нескольких файлов, связанных ключевым полем - номером рабочего по делопроизводственным документам фирмы. Реляционная база создавалась таким образом, чтобы сохранить почти все данные источника. Основной файл базы имеет статичный характер, тогда как все остальные содержат динамическую информацию. Динамический характер БД позволяет получить представление о положении любой группы рабочих в любой год и даже месяц изучаемого периода. Кодировка практически всех исходных показателей позволяет наиболее компактно хранить данные источника, агрегировать информацию тех показателей, которые имеют очень большое число мелких категорий, а также не уделять специального внимания проблемам опечаток и других ошибок источника (например, в географических названиях губерний или уездов при указании места рождения и т.п.). Наконец, кодировка и агрегирование информации делают данные пригодными для анализа с помощью многих стандартных пакетов статистической обработки.

Таким образом, динамическая модель комплекса документов первичного учета в форме максимально полной реляционной базы данных дает возможность работы с источником в "активной форме", позволяющей как извлекать информацию по запросам исследователя, так и применять различные методы анализа для решения конкретных исторических задач.

Х.А.ДИДЕРИКС, Х.Д.ТЪЯЛСМА
Лейденский университет, Голландия

КОМПЬЮТЕР В ИССЛЕДОВАНИИ ПРОФЕССИЙ

Изучение возникновения новых профессий, исчезновения одних из них и изменения смысла других может рассматриваться как специальное направление исследований, особенно в экономической и социальной истории. В области экономической исторической нас может интересовать появление новых занятий в связи с изменениями в промышленном производстве или других секторах экономики; данные о занятиях могут использоваться для экономического анализа. В исследованиях по социальной истории сведения о занятиях используются для изучения социальной структуры и наиболее интересны в качестве основного фактора, определяющего социальный статус. При этом постоянно приходится возвращаться к вопросам классификации и кодирования занятий. Вряд ли возможно разработать международную, общепринятую схему классификации занятий. Более реальным было бы сопоставление основных существующих ныне схем и

разработка программного обеспечения для связывания таких схем, что сделало бы возможными сравнение данных по разным странам.

Статья основана на анализе опыта работы по составлению базы данных по занятиям (на материале различных источников XVI-XX веков), ведущейся на факультете экономической и социальной истории Лейденского университета (Нидерланды), и "Проекта малых городов", осуществляемого в Центре по истории города при Лейсестерском университете (Великобритания).

Схема классификации занятий, используемая в Лейдене, основана на материалах переписи 1889 года. Так как индустриализация началась в Нидерландах лишь в самом конце XIX в., эта перепись отражает экономический и социальный строй традиционного общества. Поэтому эта классификация бралась за эталон почти во всех исследованиях по экономической и социальной истории Голландии и для более ранних периодов.

Классификация занятий, использовавшаяся при этой переписи, насчитывала 836 названий занятий, объединенных в 34 категории.

Занятия составляют основу социальной стратификации и для более поздних периодов, но группировка базируется на престижности отдельных занятий. Широко известная схема такого рода, выделяющая шесть статусных групп, была разработана социологом Ван Тульдером.

К настоящему времени создан ряд наборов данных, содержащих данные о занятиях, классифицированных в соответствии с переписью 1889 года. Недавно эта информация была сдана на хранение в Нидерландский Архив исторических данных. Важная часть их заимствована из Банка Данных по истории населения Лейдена, материалы которого взяты в основном из переписей населения, списков жителей и налоговых ведомостей.

Все данные о занятиях закодированы по одной схеме. Все коды уникальны; они представляют собой четырехзначные числа, одна или две первых цифры которых представляют группу занятий, а две последних - конкретное название в этой группе. Отдельно была создана база данных, содержащая коды и описания как конкретных занятий, так и их групп в соответствии с вышеупомянутой классификацией. Поскольку коды уникальны, то записи в этой базе данных легко могут быть связаны с кодами в файлах данных, перечисленных выше. Планируется дополнить каждую отдельную запись в этой базе данных, содержащей описания занятий, информацией о вхождениях соответствующего названия в файлы данных. Это позволило бы устанавливать, когда определенные названия занятий исчезают или изменяются, и в каких регионах.

База данных, содержащая названия занятий, создана в СУБД dBaseIV. Проводились эксперименты по соединению базы данных, организованной на "принципах 1889 года", с иной классификацией, используемой в "Проекте малых городов" Лейсестерского университета. Это модифицированная система Бута-Армстронга. В настоящее время ведется работа по согласованию двух схем классификации. Это означает необходимость адаптировать систему, ориентированную прежде всего на занятия (Голландская перепись 1889 г.), к системе, базирующейся на классификации отраслей промышленности (Бут-Армстронг). Категории в двух системах часто не совпадают.

Связывание двух баз данных проводится без внесения изменений в исходные файлы. Связи устанавливаются на уровне категорий при посредстве переходных файлов, содержащих информацию как о Лейденской, так и о Лейсестерской системе, и имеющих лишь по два поля: класс или группа Лейсестерской классификации и группа Лейденской системы.

Контакты с лейсестерским "Проектом малых городов" поставили, в частности, вопросы о переводе названий занятий с голландского на английский и выработке сравнимых классификаций. Мы полагаем, что в будущем появятся сравнительные исторические исследования, посвященные изучению определенных занятий населения, различиям в их названиях и изменениям этих названий во времени, в области сельского хозяйства или какой-то отрасли промышленности, например, текстильной промышленности.

Сергей ЖУРАВЛЕВ, Виктория ТЯЖЕЛЬНИКОВА
Институт российской истории РАН, Москва, Россия

ИНОСТРАНЦЫ В СССР (20-30-е годы). ПРОБЛЕМЫ ФОРМИРОВАНИЯ БАЗЫ ДАННЫХ

Поиск сведений и анализ информации об иностранных гражданах, которые приехали в Советский Союз после Октября 1917 г. и в большинстве случаев погибли в застенках ГУЛАГа, ставят целый ряд специфических проблем при создании базы данных. Круг исследовательских вопросов определяют два момента: специфика комплекса источников и постановка исследовательских проблем.

Специфика комплекса источников обусловлена:

- 1) многообразием видов письменных источников (документы личного происхождения: анкеты, личные дела, заявления, жалобы, воспоминания; делопроизводственная документация; судебно-следственные материалы);
- 2) наличием источников различных типов, в том числе бланков, печатей, кино- и фотодокументов;
- 3) двуязычным характером комплекса документов на каждого иностранца (главным образом русско-немецкий, а также русско-английский и русско-французский типы);
- 4) интернациональным характером личных сведений (например, данных об образовании, полученном в различных странах Европы, Америки, Азии в начале XX века);
- 5) своеобразным восприятием советской действительности 20-30-х годов, что отразилось в нетрадиционном употреблении советизмов и понятий эпохи.

Все перечисленные особенности комплекса диктуют необходимость решения при создании базы данных нескольких специальных задач прочтения источников. Во-первых, выявление закономерностей написания личного имени на русском языке иностранцем, русскоязычным служащим, в прессе на третьем языке. Варианты написания связаны как с элементарной небрежностью, так и с особенностями произношения и написания на слух звуков, не имеющих адекватного аналога в другом языке. От качественного формирования системы правил и закономерностей существенно зависит оперативность поиска имени в базе и выбор нужного варианта. Во-вторых, адекватная интерпретация сведений, их кодировка и расширение исследовательских возможностей обуславливают необходимость создания серии баз специальных знаний. Так, например, обработка данных об образовании невозможна без четкого уяснения ступеней среднего, специального и высшего образования в начале XX века. База знаний о политических партиях и общественных организациях позволяет в отдельных случаях определить причину эмиграции, если она была связана с конкретными политическими событиями. В-третьих, при работе с советскими источниками 20-30-х годов особую роль играют реалии этой эпохи. В данном случае они особым образом проявляются в употреблении терминов и понятий вследствие как особенностей национального восприятия, так и двуязычного характера источников. Здесь, кроме специальных баз знаний по сокращениям, аббревиатурам, значению отдельных терминов и функциональной нагрузке понятий, встает задача адекватного перевода и интерпретации слов и выражений, которые образованы вне грамматических и лексических норм и правил обоих языков (например: Starost vom Politikzirkel, Udarniker von Planerfüllung и др.)

Второй круг проблем при создании базы данных обусловлен исследовательскими задачами. Основная цель работы - разноуровневая реконструкция жизни иностранцев в СССР, поскольку при реконструкции на макроуровне доминирует тенденциозный характер источников. Особый интерес представляют следующие микроуровневые задачи: реконструкция жизни самого эмигранта; его семьи; иностранной (немецкой) колонии в отдельном городе и картины жизни в целом. Источники позволяют выявить

круг сослуживцев, соседей, знакомых, партийных товарищей, на основе чего проследить цепочки личных связей, выявить особенности отношений между иностранцами и иностранцев с русскими, а также установить каналы распространения информации с родины и ее влияние. Постановка таких задач выводит на особое место традиционные вопросы о максимальном использовании в базе т.н. дополнительной информации источника, о взаимопроверке сведений из различных документов и обработке комплекса сведений из различных баз.

Татьяна МОИСЕЕНКО

Институт Российской истории РАН

Михаил СВИЩЕВ

Московский государственный университет им.Ломоносова, Россия

БАЗЫ ДАННЫХ ПО НОВЕЙШЕЙ ИСТОРИИ СЕЛЬСКОГО ХОЗЯЙСТВА И КРЕСТЬЯНСТВА РОССИИ (к проблеме вторичного использования)

Доклад посвящен проблеме вторичного использования баз данных по истории сельского хозяйства и крестьянства России: условиям доступа к информации, возможностям обмена информацией между различными группами пользователей, стандартизации данных и др.

Создание баз данных - важная задача, но не менее важна и проблема их эффективного использования в научных исследованиях. Однако зачастую бывает легче вновь обратиться к источникам, чем получить доступ к чужим базам. Так, согласно статистике Банка социологических данных (Москва), более половины пользователей информационных баз по социологии - это сами их создатели. Можно с уверенностью сказать, что в области истории положение значительно хуже.

Особые сложности возникают при попытках использовать в исторических исследованиях информационные массивы, созданные вне рамок научного (исторического) сообщества. В данном случае речь пойдет о базах и коллекциях машиночитаемых данных, которые были созданы различного рода государственными и общественными институтами и предназначались для решения практических задач управления. В последнее время крупные базы данных, содержащие разнообразные характеристики социального и экономического развития российской деревни, стали создаваться также различными коммерческими структурами.

Сейчас МЧД по истории сельского хозяйства и крестьянства, отложившиеся в результате деятельности различных структур, по существу не используются историками, несмотря на то, что они нередко содержат уникальную информацию. Историки крестьянства зачастую не знают не только о структуре и содержании, но даже о существовании подобных информационных массивов. Не меньшую трудность представляют различия в подходах к отбору и моделированию данных, predeterminedенные разными целями создателей и пользователей баз данных, а также использование не всегда совместимого программного обеспечения. Доступ к информационным ресурсам государственных и коммерческих структур существенно ограничен отсутствием нормативных актов, регулирующих их информационную деятельность, а также высокой стоимостью сервисных услуг, что делает такие ресурсы недоступными не только для отдельных исследователей, но и для научных институтов. В настоящее время ситуация значительно осложняется еще и потому, что после "отмирания" многих государственных и общественных структур бывшего СССР над их информационными массивами нависла угроза уничтожения.

Проиллюстрируем эти положения на примере нескольких баз данных, различных по своему происхождению и характеру содержащейся в них информации.

Самой крупной базой данных о российской деревне последнего десятилетия является база, созданная на основе годовых отчетов колхозов и совхозов, общее число которых в конце 80-х годов превышало 50 тыс. Она стала формироваться в 1978 г. после избрания М.С.Горбачева секретарем ЦК КПСС по сельскому хозяйству. Цель создания базы заключалась в мобилизации информации для принятия решений отделами и секторами ЦК, оперативного управления сельскохозяйственным производством вплоть до уровня райкомов, составления аналитических и прогностических материалов. Работа по сбору информации до 1988 года велась специальным подразделением в составе сельхозотдела ЦК при участии ЦСУ, после чего в связи с реорганизацией аппарата ЦК пополнение базы прекратилось. В базе данных содержатся важнейшие показатели экономического и социально-демографического характера: данные о динамике и объеме сельскохозяйственного производства по отраслям, технической вооруженности труда, составе кадров специалистов и работников сельского хозяйства, обеспеченности их жильем, а также сведения о состоянии личного подсобного хозяйства, т.е. единственно допускаясь в СССР формы мелкого частного производства. Основная черта базы - ее информационная избыточность. Однако несмотря на огромные затраты на ввод данных и ценность собранной информации, она использовалась мало; потенциальные пользователи, предпочитавшие готовые справочные материалы, даже не подозревали о возможностях системы. Кроме того, база данных имеет узко-специализированное программное обеспечение, что затрудняет доступ к ее информации.

На иных принципах создавалась база данных, в основе которой лежат более 70 тыс. заявлений граждан, желающих переехать в сельскую местность. Они поступили в ответ на публикацию объявления в 4 крупнейших российских газетах, тираж которых превышает 20 млн. экз. Перевод информации в машиночитаемую форму и ее обработка ведется неправительственным агентством, которое занимается оказанием помощи русским беженцам из национальных регионов бывшего СССР, и финансируется предпринимательскими структурами. В базу заносится информация, характеризующая потенциального переселенца и его семью: возраст, образование, профессия, специальность, фиксируется также, чем собирается заниматься в деревне будущий фермер. Программа обработки данных ориентирована на поиск определенного объекта по узкому набору признаков, имеющих практическое значение с точки зрения организации процесса переселения. Информация предоставляется только на коммерческих началах, независимо от характера использования.

Еще одним вариантом баз данных, которые создаются для решения практических задач, но могут обрести вторую жизнь при использовании их информации историками, являются базы, содержащие материалы социологических исследований. Они предназначены, как правило, для определенных заказчиков, в качестве которых выступают различного рода государственные и общественные структуры и средства массовой информации, и используются в течение нескольких месяцев, а иногда и нескольких дней. Проблема их вторичного использования широко обсуждается в литературе, но до сих пор далека от своего решения. Примером может служить база, созданная на основе опросов сельского населения на территории России, Украины и Белоруссии (выборка объемом более 2,5 тыс. чел.), проведенных Всесоюзным (ныне Всероссийским) центром изучения общественного мнения в 1990 и 1992 гг. по заказу журнала "Сельская молодежь". Анкета насчитывает более 100 вопросов, характеризующих отношение сельского населения к аграрной и земельной реформе, перспективам развития частной собственности на землю, создания фермерских хозяйств. Материалы данного опроса представляют уникальный материал для изучения общественного сознания крестьянства. Программным обеспечением базы является модифицированная версия FoxBase, на основе которой был разработан Комплекс Анализа Опросов Населения (КАНОН), позволяющий проводить обработку и репрезентацию данных в форме, необходимой для публикации в социологическом

бюллетене ВЦИОМ. Наличие специального программного обеспечения существенно сужает круг потенциальных пользователей информации. Использование базы данных, которая в настоящее время хранится в ВЦИОМ, хотя может быть уничтожена, возможно как на коммерческой основе, так и на началах научного обмена.

Альберт МЮЛЛЕР

Институт историко-социальных исследований им. Людвига Больцмана Вена, Австрия

МИГРАЦИЯ И ПРОБЛЕМА ПОДВИЖНОСТИ НАСЕЛЕНИЯ В СРЕДНИЕ ВЕКА (НА ОСНОВЕ ИЗУЧЕНИЯ "ПОСЕТИТЕЛЕЙ" ВЕНСКОГО УНИВЕРСИТЕТА)

Поскольку европейские государства рассматривают увеличение миграционных процессов и географической мобильности населения, как современную проблему, постольку постоянно возрастает широкий общественный интерес к изучению истории миграций и мобильности. В этой связи чрезвычайно важным и заслуживающим внимания представляется изучение миграционных процессов нового и новейшего времени. Некоторые исследования, выполненные в этой области как на международном, так и на местном (локальном) уровне, помогли до некоторой степени изменить, уточнить и частично пересмотреть устоявшееся представление о традиционном доиндустриальном обществе как об обществе статичном и не подверженном миграционной мобильности. Проводящееся в настоящее время исследование миграции в период средневековья и доиндустриальной эпохи демонстрирует тесную связь этих процессов со специфическими аспектами, присущими исследуемым периодам. Очевидно, что принадлежность к определенной социальной группе "предсказывает" (причем, весьма точно) "миграционное поведение" лиц, составляющих эту группу.

Одной из наиболее интересных социальных групп является группа под условным названием "посетители" средневековых университетов, которая включает в себя студентов, сотрудников, преподавателей университетов. На примере Венского университета можно показать, что география тех мест, откуда "посетители" приезжают в университет, охватывает Россию на Востоке, Шотландию на Западе, Италию на Юге и Финляндию и Швецию на Севере. Однако, наиболее многочисленно и полно представлены области германоязычных стран и Венгрии, а также некоторых славянских государств.

Данные тезисы предлагают для обсуждения некоторые вопросы, возникшие при создании базы данных по биографическим данным слушателей средневекового Венского университета. Источниками для базы данных послужили списки зачисленных в Венский Университет, в которых содержатся сведения о 55 тыс. персон за период с 1377 по 1554 гг. Опубликованное издание названного источника было преобразовано в машиночитаемую форму с помощью пакета KDEM. Данные списки представляют собой довольно простой структурированный текстовый источник. "Посетители" Университета упоминаются в порядке их прибытия и в соответствии с их принадлежностью к одному из 4 землячеств. Эта присущая источнику структура очевидно диктует иерархическую модель данных.

Количество информации, относящейся к каждой отдельной личности, может быть различным. Естественно, для каждого человека записывалось его имя. Для 96% изучаемых лиц указано географическое место происхождения. Особо фиксировались такие параметры, как "дворянский титул", "ученая степень" и т.д.

Машиночитаемый источник обрабатывался при помощи пакета программ "StanFER", обладающего высокой степенью "чувствительности" к контексту и позволяющего выделить семантические "сущности" источника и организовать его оригинальный текст в соответствии со структурой модели данных.

Кроме этого, база данных была обработана с помощью системы КЛИО. Широкие возможности этого уже хорошо известного пакета программ для историков предназначены для выполнения большого числа оценивающих процедур. После проведения кодировки появляется возможность получить из тестовой базы количественные данные, полностью подготовленные для статистического анализа. Некоторые результаты этой работы будут представлены на конференции. В соответствии с афоризмом Я.Вергера "от статистики к просопографии", предстоящая работа будет сконцентрирована на различных возможностях "сведения" воедино различных источников. Университетский Архив в Вене содержит огромное число неопубликованных документов со сведениями об университетских курсах и данными о дальнейшей карьере университетских "посетителей". Собранные вместе, все эти источники создают разновидность коллективной биографии, позволяющей связывать записи разных источников, что обеспечивается возможностями пакета КЛИО. Этот важный этап работы находится пока на стадии поиска, тестирования и апробирования оптимальных принципов для объединения данных различных источников.

Алексей ТИХОНОВ

Московский государственный университет им.Ломоносова, Россия

СОЗДАНИЕ БАЗЫ МАШИНОЧИТАЕМЫХ ДАННЫХ НА ОСНОВЕ КОМПЛЕКСА РАЗНОТИПНЫХ ИСТОЧНИКОВ

(Бельгийские компании в России конца XIX - начала XX века)

Главной трудностью для историков, изучающих проблемы частного предпринимательства в России, всегда оставалась недостаточная источниковая база для проведения таких исследований (в частности, плохая сохранность фондов частных компаний). В нашей работе мы попытались определить возможные пути решения этой проблемы за счет нового прочтения уже введенных в исторический оборот источников. Таковыми являются справочно-информационные издания по торгово-промышленным компаниям, вышедшие в России в начале века, и дела об их учреждении, хранящиеся в фондах министерств финансов и торговли и промышленности.

Так как материалы справочников представлены по определенной, жестко заданной структуре, это позволяет перевести их в машиночитаемый вид без каких-либо существенных потерь информации. При этом был выработан единый формуляр записи сведений по каждой фирме, который может использоваться и в дальнейшем, для материалов других изданий. Сведения об учредителях, имеющиеся в уставах акционерных компаний, также представляют собой высокоструктурированный источник. Объектом нашего исследования стали компании, действовавшие в России в конце XIX - начале XX вв. на основе бельгийских уставов. В результате была создана реляционная база машиночитаемых данных RUBELGE, которая представляет собой композицию трех основных файлов: COMP_1903, COMP_1914, ADMIN.

В файле базы данных по 1903 г. (COMP_1903) имеется 122 записи, что соответствует количеству бельгийских компаний, действовавших в России в это время по сведениям "Указателя действующих в империи акционерных предприятий" (Спб., 1903).

В файле COMP_1914 находятся 72 записи, соответствующие количеству бельгийских компаний по материалам изданного в 1914 г. "Сборника сведений о действующих в России акционерных обществах и товариществах на паях" (Спб., 1914).

Оба файла, описывающие компании, имеют одинаковую структуру из 24 полей, содержащих следующую основную информацию о действовавших компаниях.

1. Название на русском и французском языках;
2. Отрасль;
3. Предприятие;

4. Место действия;
5. Дата основания в Бельгии;
6. Дата разрешения действий в России;
7. Дата открытия ответственного агентства (фактическое начало действий);
8. Имя ответственного агента;
9. Местопробывание ответственного агентства;
10. Местопробывание администрации;
11. Количество, тип и номинал акций (а также наличие учредительских паев и других типов ценных бумаг);
12. Размер уставного капитала;
13. Размер прибылей или убытков;
14. Размер кредитных обязательств;
15. Дополнительная информация.

Третий файл, ADMIN, отражает состав учредителей этих предприятий, которые входили в Совет правления каждой фирмы, избранный на первый срок. Файл состоит из 405 записей. Каждая запись соответствует одному администратору. "Карточка" учредителя имеет следующий вид:

1. Порядковый номер;
2. Порядковые номера компании по справочникам;
3. Имя учредителя;
4. Профессия;
5. Адрес;
6. Количество и номинал акций;
7. Наличие учредительских паев (и их количество);
8. "Представительство" другой компании через члена правления.

Количество записей больше, чем число реальных людей (314), т.к. некоторые предприниматели могли входить в правления сразу нескольких бельгийских компаний, действовавших в России.

Главной целью автора работы было переведение сведений из источника в машиночитаемую форму без потерь каких-либо данных. Речь идет именно об исходных вариантах баз, так как в процессе работы возможно строить любые производные файлы (путем агрегирования, кодировки и т.д.) для постановки конкретных исследовательских задач.

Созданная таким образом реляционная база данных, представляющая собой объединение нескольких источников, открыта для новой информации и может служить исходным материалом для анализа с помощью таких статистических пакетов, как STATGRAPH, SPSS, и др.

Кевин ШУРЕР

Кембриджская группа ESRC, Кембриджский университет, Великобритания

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ: ВЗГЛЯД ИЗ ВЕЛИКОБРИТАНИИ

Внедрение компьютерной технологии в исторические исследования, по мнению многих исследователей, оказало революционизирующее влияние на историю. Попытаемся оценить воздействие проникновения информационной технологии в историческую науку на опыте Великобритании. Рассмотрим ситуацию за последние 15 лет, в особенности в области перехода от ориентации на большие ЭВМ к использованию микрокомпьютеров.

В этом направлении попытаемся описать цели деятельности английской Ассоциации History & Computing, основанной в 1986 г. Ее деятельность на научном уровне сконцентрирована на организации конференций, повышении квалификации историков путем организации летних школ и на публикации монографий и периодических изданий. Все эти формы способствуют созданию условий для обмена идеями, опытом и результатами исследований, обеспечивая тесные контакты прежде разрозненной группы ученых. На более высоком, международном уровне Ассоциация предлагает к обсуждению на специальных семинарах ключевые вопросы, касающиеся развития компьютерных технологий в истории. Такой подход позволяет не только выявить основные теоретические и методические проблемы, но и укрепляет репутацию нашей Ассоциации как лидирующей в этом направлении.

Однако, расширение применения компьютерных технологий в исторических исследованиях выявило не только достоинства этого направления, но и множество проблем. Теоретически, сохраняется опасность, что развитие этих технологий могло бы не сблизить, а скорее увеличить разрыв между теми, кого можно отнести к "традиционалистам" и "технократам". Если это так, то, по-видимому, требуется оценить, каким образом компьютерные технологии должны внедряться в исследовательский процесс. Помимо этого, у историков, использующих эти технологии, есть и собственные, "внутренние" проблемы, требующие решения. Часть из них - стандартизации кодирования, создание программного обеспечения, сбор и хранение данных и внедрение обобщенных языков описания данных - рассматриваются на семинарах и конференциях АНС и еще многократно будут обсуждаться. Наиболее актуальными из этих проблем нам кажутся проблемы адекватного хранения, обмена и распространения информации. Исходя из этого, в докладе будет охарактеризован опыт недавно завершеного проекта, позволяющий оценить возможность создания архива исторической информации в Великобритании, будут приведены результаты и выводы этого проекта, а также перспективы будущей работы.

Анализ хозяйства и семьи в исторической перспективе:

старые идеи и новые методы

Изучение истории хозяйства и семьи получило развитие прежде всего в последние 20 лет и в некотором смысле может рассматриваться параллельно с развитием компьютерных технологий в истории. Рассмотрим два сюжета. Первый касается исследовательского проекта Кембриджской группы ESRC по истории населения и социальных структур, который изучает структуры хозяйства и семьи в Англии и Уэльсе в конце XIX-начале XX вв. Второй сюжет касается методической стороны изучения семей и хозяйств этого периода и показывает необходимость нового подхода.

Конец XIX-начало XX вв. в Англии и Уэльсе - это период демографических изменений. Уровень фертильности в первых двух десятилетиях XX в. снижался при одновременном резком снижении младенческой и детской смертности. В результате этот период знаменует начало того, что было названо долговременным сдвигом возрастной структуры - это изменение возрастной структуры населения в сторону более старых групп. По-видимому, в структуре хозяйств этого периода есть некоторые изменения, фундаментальные для понимания перехода от "традиционной" к "современной" семье. Но недостаток необходимых данных переписей затрудняет детальное изучение структуры хозяйств этого периода. К счастью, Бюро переписей населения имеет выборку данных по переписям 1891, 1901, 1911 и 1921 гг., содержащую почти 0.5 млн персональных записей. В докладе анализируются некоторые проблемы обработки этого массива данных, а также предлагаются некоторые предварительные результаты проекта. Второй сюжет концентрируется на методической стороне исследования истории хозяйства и семьи. Он подчеркивает, что в методическом плане уровень анализа несколько выходит за пределы того, который продемонстрирован в плодотворной работе

Ласлетта по этой теме (Хозяйство и семья в прошлом; издание 1972 г.). В заключение предлагается альтернатива стандартной формы классификации структуры хозяйств, предложенной Ласлеттом, на уровне отдельных индивидуумов. Этот новый подход, как будет продемонстрировано в докладе, имеет большую степень гибкости по сравнению с подходом Ласлетта, не только благодаря тому, что он основан на изучении именно индивидуумов, а не целых хозяйств, но также на больших возможностях статистического анализа такой информации.

Юлия ЮМАШЕВА

Московский государственный университет им Ломоносова, Москва, Россия

ИСТОЧНИКОВЕДЧЕСКИЕ ПРОБЛЕМЫ СОЗДАНИЯ ПРОСОПОГРАФИЧЕСКИХ БАЗ ДАННЫХ

Проблема создания баз данных по биографическим характеристикам отдельных личностей и лиц, входящих в целые социальные группы (просопографических баз данных) занимает многих историков. Научно-исследовательской целью создания просопографической базы данных, как правило, является возможность формирования так называемой "коллективной биографии" исследуемого социума.

Одним из важнейших методологических аспектов проблемы является вопрос о соотношении терминов, употребляемых в этой области. Термины "просопография" и "коллективная биография", характерные для зарубежных исследователей, отражают особый тип исследования, подразумевающий изучение жизненных путей (динамических показателей) отдельных людей (существует прекрасный русский эквивалент этого слова - "жизнеописание", который очень ярко отражает существо процесса). Однако в отечественной историографии превалирует понятие "облик", который, по нашему убеждению, совершенно определенно характеризует конечный результат подобных исследований - получение статичного "коллективного портрета" исследуемой социальной группы. Это различие, возможно, коренится в различных типах используемых источников и различных целях исследования.

Как правило, большинство работ, проведенных в рамках создания банков данных биографической информации, базируется на "классических" видах массовых источников структурированного вида, т.к. массовые источники (в основном, первоисточники) в виде единых формуляров наиболее удобны для количественного анализа.

Остановимся на проблемах, которые возникают при работе со структурированными источниками, и рассмотрим несколько "рабочих" вариантов.

ВАРИАНТ 1. 1-1 ("Один источник - одна персоналия")

Используется массив структурированных документов, в котором каждый источник единичен и содержит сведения об одном лице (личные дела, анкеты и т.д.). В этом случае снимается вопрос о достоверности каждого отдельно взятого документа, и возможно создание выборки как среди источников, так и среди персоналий (своего рода "безличностный" подход). На основе этих источников получается "коллективный портрет", т.к. они, как правило, отражают одномоментное состояние объекта изучения.

ВАРИАНТ 2. N-1 ("Несколько источников - одна персоналия")

К исследованию привлекаются несколько однотипных источников об одном и том же лице (например, анкета + автобиография + личное дело + . . . + и т.д.). Источники сравнимы и легко формализуемы, но при этом они различны и формирование выборки невозможно. Однако появляется вероятность дополнения сведений одного источника информацией другого, а также "перекрестного" анализа и проверки сведений об одних и тех же людях, содержащихся в разных источниках, с выявлением разночтений, неточностей и т.д. В результате работы получается все тот же "коллективный портрет",

хотя вариант имеет потенциальную возможность (при особом исследовательском подходе) "перерасти" в "коллективную биографию".

ВАРИАНТ 3. 1-N ("Один источник - несколько персоналий")

Этот вариант возможен, когда в качестве источника для исследования и формирования базы данных биографического характера служат различные справочники, биографические словари, продолжающиеся издания, в которых биографии интересующих лиц описаны по одной общей схеме.

Следует заметить, что само понимание термина "массовый исторический источник" в приведенных рассуждениях несколько сужено и сведено к кругу материалов, которые на практике удовлетворяют следующим требованиям: 1) единообразная форма; 2) единообразное содержание; 3) большое количество документов. На наш взгляд, эта трактовка не совсем верна, и ее следует расширить за счет более широкого привлечения к просопографическому исследованию источников нарративного характера, которые более многочисленны, но менее пригодны для машинной обработки. Рассматриваемый случай соответствует, вероятно, варианту 4, который можно представить в виде формулы N-N ("Несколько источников - несколько персоналий"). Особенности нарративных источников является их уникальность, что делает невозможным выборку, трудность формализации (т.е. создания формуляра и макета кодировки), субъективность содержания (например, в художественных произведениях). Наконец, они разнятся по типам, жанрам, целям создания и по количеству сведений для каждой отдельной личности, что значительно усложняет работу по созданию единого формуляра и ведет к необходимости первоначального формирования сводного метаисточника. Кроме того, следует особо выделить проблему сохранения "внешних" характеристик (реквизитов) источников, которая также обостряется в случае работы с разнотипными документами. При разработке метаисточника возникает множество проблем (рассматривается вариант "Несколько разнотипных нарративных источников - одна персоналия"):

- информация в источниках "пересекается" и совпадает;
- существуют разночтения в разных источниках как непротиворечивые (из-за разной детальности изложения), так и полностью взаимоисключающие;
- данные в источниках неполны по разным показателям;
- данные избыточны;
- информация носит отрывочный характер - сведения практически полностью отсутствуют (единичные упоминания).

Опыт автора по созданию источника такого рода по данным о высшем генералитете советской армии периода Второй мировой войны свидетельствует как о возможности решения подобных проблем в конкретном историческом исследовании, так и необходимости углубленной теоретической разработки источниковедческих проблем, связанных с созданием просопографических баз данных на материалах разнотипных источников.

Питер ДООРН

Лейденский университет, Голландия

ГЕОГРАФИЧЕСКОЕ ПОЛОЖЕНИЕ, МОДЕЛИ ВЗАИМОДЕЙСТВИЯ И РЕКОНСТРУКЦИЯ ИСТОРИЧЕСКИХ ПОСЕЛЕНИЙ И КОММУНИКАЦИЙ (на примере Этолии, центральная Греция)

Хотя географию и историю можно рассматривать как родственные дисциплины, историки редко используют пространственные модели для изучения прошлого. В данной работе делается попытка показать, как информация географического характера может

применяться для реконструкции локализации исторических поселений и коммуникаций между ними.

В течение прошлого десятилетия появилось специальное программное обеспечение, ориентированное на хранение, анализ и представление картографической информации (т.н. GIS - географические информационные системы). Обычно эти системы анализируют картографические данные в сочетании с дополнительной информацией; как правило, GIS обладают возможностями компьютерного картографирования (mapping programmes), статистических пакетов и программного обеспечения баз данных. Следует различать растровые процедуры, в которых карта рассматривается в форме сетки с регулярными "ячейками"-квадратами и векторный подход, основанный на различных представлениях таких структурных единиц, как точки, линии и элементы поверхности. Целью междисциплинарного исследовательского проекта Aetolian Studies Project, предпринятого с начала 1980-х годов, была реконструкция истории заселения этой исторической области Центральной Греции - с доисторической эпохи до нашего времени. Информация о расположении поселений, их природном окружении и взаимоотношениях этих поселений была собрана из широкого круга источников: археологических, церковных, налоговой документации времен Османской империи, исторических и современных топографических карт, материалов переписей и статистических данных, счетов на транспортные расходы и военных справочников, данных аэрофотосъемки, монографических исследований, материалов опросов сельского населения и полевых исследований. Большая часть этой информации была занесена в базы данных и текстовые базы.

В данной работе для анализа указанной информации используется ряд аналитических подходов, развитых в исследованиях по "human geography":

- модели расселения: для изучения сдвигов в расположении поселений во времени;
- гравитационные модели и модели потенциалов: для анализа возможных коммуникаций между поселениями;
- аналитические средства обработки топографических данных и информации о рельефе: для установления исторических торговых путей.

1. Модели расселения.

Хотя общая теория размещения поселений пока не создана, существует несколько классических моделей для количественного описания и объяснения первичной, вторичной и "третичной" деятельности: теория земельной ренты (Рикардо и др.), имеющая дело с размещением сельскохозяйственных "единиц"; теория пространственной локализации промышленности, в основном базирующаяся на работах Альфреда Вебера; теория "центральных мест" (Кристаллер и др.), применяемая для описания размещения коммерческих предприятий и сервиса.

Анализ литературы показывает, что наиболее важными факторами локализации поселений являются: доступность земельных и водных ресурсов, коммуникации, безопасность, микроклимат, а также фактор культуры. Конечно, присутствует и элемент случайности.

Чизхолм описал процесс локализации сельских поселений как результат воздействия ряда факторов, "подводящих" деревню к определенному местоположению; при этом каждый фактор описывается как функция от расстояний и стоимостей. Можно расширить эту модель для анализа региона в целом, используя методы GIS. Анализируя определенные районы Этолии в рамках данного подхода, мы использовали метод "квадратных ячеек". На основе топографических карт (масштабом 1:50000) были введены наиболее важные факторы локализации: высота и наклон местности, ее описание, землепользование и водные ресурсы, а также исторические торговые пути. Наряду с этим использовалась информация обо всех поселениях, когда-либо зафиксированных в данной местности или существующих в данное время. В результате появляется возможность анализировать с помощью различных методов соотношения

между расположением поселений и факторами локализации. Связи между отдельными факторами и выбором места поселения можно анализировать в динамике, выявляя, например, изменения роли таких факторов, как безопасность, земельные и водные ресурсы, коммуникации. Возможно также проводить группировку различных факторов локализации, соотнося их с "лучшими" местами для поселения. Варьируя на базе теоретических построений веса различных факторов, можно показать, как под воздействием исторических изменений менялась структура предпочтений при выборе места поселения.

2. Гравитационные модели.

Гравитационная модель в теории поселений вводится по аналогии с известной формулой из физики, согласно которой сила притяжения двух физических тел прямо пропорциональна произведению их масс и обратно пропорциональна квадрату расстояния между ними. При использовании этой модели в географии ньютоновский закон притяжения трактуется менее строго; модель предполагает, что при прочих равных факторах интенсивность коммуникаций между двумя поселениями прямо пропорциональна их размерам и обратно пропорциональна расстоянию между ними. Агрегирование взаимодействий каждого поселения со всеми остальными из рассматриваемой местности позволяет определить его "коммуникационный потенциал". Таким образом, зная численность населения в каждом из поселений и расстояния между ними, мы можем вычислить потенциалы коммуникаций в данном районе и оценить влияние на них со стороны изменений в распределении населения. Применение методов кластер-анализа к матрице интенсивностей парных коммуникаций позволяет выделить группы взаимосвязанных поселений или "узловые" регионы.

3. Реконструкция торговых путей.

Хотя гравитационные модели дают интересное представление о коммуникационных потоках и об агрегированной структуре множества изучаемых поселений, этот подход все же недостаточен для реконструкции реально существовавших в прошлом торговых путей. Здесь мы можем использовать другую модель, заимствованную также из естественных наук и основанную на законе сохранения энергии. С помощью этой модели можно реконструировать наиболее вероятные торговые пути, исходя из данных о коммуникационных центрах и концентрации населения.

Основное модельное предположение сводится к тому, что человек стремится минимизировать усилия при достижении своих целей (в данном случае - для того, чтобы добраться в определенный пункт). На плоскости кратчайший путь между двумя пунктами пролегает по прямой линии, однако сложный ландшафт Этолии, высокие горы и глубокие долины создают естественные преграды для путей сообщений. Введя в GIS количественные данные о характере ландшафта по всей площади рассматриваемой местности (высоты и характеристики наклона гор, русла рек), мы можем получить оценку количества энергии, требуемой для преодоления пути из пункта А в пункт В, в виде произведения "суммарной сложности" определенного маршрута и расстояния. Вычисляя энергию, потребную для альтернативного пути, мы по сути моделируем попытки поиска "оптимального" пути, имевшие место в исторической реальности. При этом естественно предполагать, что наиболее вероятным был в прошлом тот путь, которому соответствовали минимальные затраты энергии.

В докладе приводятся конкретные результаты применения рассмотренных моделей в рамках исследовательского проекта Aetolian Studies Project.

Ян ОЛДЕРВОЛЛ

Бергенский университет, Норвегия

WINCENS ИЛИ НОВАЯ СИСТЕМА ДЛЯ РАБОТЫ С ПЕРЕПИСЯМИ

Историки постоянно ищут универсальные средства для обработки данных, но это невыполнимая задача. Можно лишь пытаться создавать средства, которые были бы лучшими в определенном смысле. Работая со структурированными историческими источниками и создавая программное обеспечение для этой работы на новом уровне компьютерной технологии 90-х гг., автор базируется на уже существующей системе CensSys.

Система CensSys создавалась автором в течение нескольких последних лет и представляет собой довольно простую СУБД (систему управления базами данных). Не обладая некоторыми свойствами, типичными для СУБД, эта система, однако, имеет и такие возможности, которые не совсем обычны для систем управления базами данных, например, возможность сжатия информации, возможность более свободного поиска в базе или возможность смешения информации различных типов в одном и том же поле. Помимо свойств СУБД, CensSys обладает и некоторыми возможностями статистических пакетов, например, строит таблицы сопряженности.

Все операции система выполняет с очень высокой скоростью, значительно превышающей скорость работы СУБД или статистических пакетов, а возможность активно связывать статистические таблицы с исходными данными позволяет при нажатии одной клавиши немедленно увидеть материал переписи, "стоящий" за каждой цифрой таблицы сопряженности. Кроме того, системой чрезвычайно легко пользоваться, она обладает многоуровневым меню, контекстно-чувствительной помощью и встроенным редактором.

Разумеется, CensSys имеет и определенные ограничения: ее структура является довольно жесткой (два и только два уровня иерархии, максимум по 50 переменных на каждом уровне), объекты в уже созданной базе практически нельзя удалять или добавлять. В каких же направлениях возможны усовершенствования?

Можно, разумеется, убирать многие ограничения CensSys: на длину поля, число переменных и объектов т.д. Более сложным является вопрос с уровнями иерархии, т.к. введение, например, пяти уровней вместо двух не является кардинальным решением проблемы и система все равно останется достаточно "жесткой". Эти соображения, а также наличие современных средств работы на персональных компьютерах, таких, как программные оболочки типа Windows, привели автора к мысли о создании другой, новой системы, получившей имя Wincens (от Windows CensSys, т.е. CensSys для работы в Windows). Завершение новой системы планируется к лету 1994 г., хотя первая ее версия уже готова.

Новая система совершенно отказывается от концепции уровней иерархии, учитывая, что многообразие связей в некоторых задачах, например, восстановления истории семей, часто разрушает строгую иерархичность данных. Вместо иерархии вводится более нейтральное понятие связи: индивидуума - с домохозяйством, домохозяйства - с домом, дома - с церковным приходом и т.д.. Таким образом можно строить реальные иерархические системы, но и не только их. Каждый индивидуум может быть связан, например, с местом своего рождения, со своим отцом, живущим где-то в другом месте и т.д.. В результате один и тот же человек может входить в несколько иерархических систем и в каждой из них может находиться на разных уровнях иерархии. Система Wincens состоит из связанных таблиц, напоминая реляционную СУБД, но она не ограничена реляционной моделью данных, т.к. связи между таблицами осуществляются не по отдельным полям, а по более гибкой схеме.

Таким образом, Wincens обладает достоинствами реляционных СУБД, не наследуя их недостатков: система гораздо более компактна в хранении информации и обладает очень высокой скоростью представления и обработки данных.

Обсуждая перспективы развития программного обеспечения для историков в более широком плане, совершенно необходимо планировать кооперацию отдельных

исследователей для обмена как идеями, так и стандартизованными программными продуктами (обработка имен, обработка дат, мер и т.п.).

Елена ОСОКИНА
Институт российской истории РАН
Георгий САТАРОВ
Центр ИНДЕМ, Москва, Россия

ПРИКЛАДНЫЕ ВОЗМОЖНОСТИ СТАТИСТИЧЕСКОГО ПАКЕТА "CLAMS" (АНАЛИЗ РАСПРЕДЕЛИТЕЛЬНОЙ ПОЛИТИКИ СОВЕТСКОГО ПРАВИТЕЛЬСТВА В ПЕРВОЙ ПОЛОВИНЕ 30-Х ГОДОВ)

1. Статистический пакет "CLAMS" разработан Центром прикладных политических исследований ИНДЕМ и творческим коллективом "Статистика". Он включает методы, которые относятся к новому разделу математической статистики - анализу данных. Кроме известных методов (множественная регрессия, метод главных компонент и др.) в пакете реализованы уникальные методы анализа данных (метод построения аддитивных кластеров, алгоритмы построения размытых классификаций, методы многомерного и индивидуального шкалирования и др.). Система "CLAMS" предназначена для использования на персональных компьютерах, совместимых с IBM PC/XT или AT. Рассчитана на пользователей, не являющихся профессионалами в статистике и программировании. В данном исследовании был использован метод метрического многомерного шкалирования.
2. Источниковой базой исследования стали архивные данные Наркомата торговли о централизованном распределении в первой половине 30-х гг. наиболее важных продуктов питания и промтоваров (два десятка наименований). На основе сведений о распределении этих товаров и данных ЦУНХУ о численности населения были получены показатели городских и сельских душевых поставок в области, края, автономные образования РСФСР, союзные республики в 1931-1933 гг.
3. Цель исследования - выявление факторов, определяющих политику централизованного снабжения населения в период карточной системы 1928-1935 гг. Статистической обработке предшествовало построение теоретической модели, объясняющей разнообразие исходных данных. Первое: различное снабжение разных территорий в конкретном году определялось разным статусом городов и аграрных районов с точки зрения проводимой Центром политики снабжения. Второе: различие снабжения могло иметь место не только между территориями в конкретном году, но могло происходить и от года к году. Это могло объясняться как изменением свойств конкретных территорий, так и изменением приоритетов самой распределительной политики. Собственно, политика снабжения и проявляется в том, какие именно свойства территорий учитываются и каким образом.
4. Математико-статистической обработке предшествовала процедура стандартизации душевых показателей снабжения. Это позволило перейти от данных, измеренных в разных шкалах (рубли, тонны), к соизмеримым величинам. Кроме того, сравнивая территории между собой по характеру снабжения, важно учитывать их различие не просто по снабжению тем или другим товаром (это можно сделать и визуально), а важно учесть в этих различиях снабжение всеми товарами одновременно. Для этого по формуле евклидова расстояния были вычислены величины, которые агрегировали, обобщали различия между парами территорий по снабжению всеми товарами.
5. Для анализа преобразованных матриц была использована модель "индивидуального многомерного шкалирования". Мы предположили, что существует некоторое число свойств, учитываемых распределительной политикой. Каждой территории эти свойства присущи в разной мере. Существует некоторое неизвестное число, которое

характеризует степень проявления свойства для данной территории. Обозначим его Z_{ij} (i -индекс территории, j -индекс свойства). В соответствии с моделью индивидуального шкалирования различия в распределительной политике проявляются в том, что в различные годы разные свойства могут быть то более, то менее важны при осуществлении распределения товаров. Степень важности того или иного свойства в каждый из годов может задаваться некоторой неотрицательной величиной W_{kj} (k -индекс года, j -индекс свойства). Фундаментальное предположение модели индивидуального шкалирования состоит в том, что вычисленные нами по исходным данным показатели различия территорий в снабжении товарами могут быть объяснены значениями свойств Z_{ij} и весов W_{kj} . Таким образом, по известным значениям различия территорий в снабжении товарами мы можем определить неизвестные значения Z_{ij} (степень проявления для данной территории свойств, учитываемых политикой) и W_{kj} (степень важности этих свойств в распределительной политике разных лет). Метод индивидуального шкалирования позволяет находить такие значения Z_{ij} и W_{kj} , чтобы ошибка была наименьшей. Для этого использовался алгоритм канонической декомпозиции Эккарта - Янга.

6. В итоге применения метода индивидуального многомерного шкалирования исследователь получает два типа таблиц. Первая содержит значения Z_{ij} . Столбцы этой таблицы представляют некие свойства территорий, которые учитывались распределительной политикой. Для того, чтобы понять, какие именно свойства территорий были выделены, необходимо по каждому столбцу посмотреть, какая группа территорий получает наибольшие и наименьшие значения. Результаты анализа необходимо сопоставить с известными исследователю дополнительными сведениями об этих группах территорий. Эту задачу можно уподобить интерпретации факторных нагрузок в факторном анализе. Интерпретируя смысл этих свойств, интересно привлекать значения показателей снабжения территорий различными товарами. Это дает возможность понять, как именно учитывались выявленные нами свойства территорий в распределительной политике Центра (например, вели к увеличению или уменьшению снабжения). После интерпретации выявленных свойств можно переходить к анализу второй таблицы. Ее значения (W_{kj}) показывают степень важности выявленных свойств в распределительной политике разных лет, а также степень их важности для городского и сельского снабжения. Чем выше значения весов в таблице, тем большее значение имело соответствующее свойство территорий в распределительной политике Центра.

7. В результате исследования было выявлено, что централизованное городское и сельское снабжение в 1931-1933 гг. определялось различными факторами. Действие этих факторов было стабильно на протяжении указанного периода. Для городского снабжения главное значение имел промышленный облик городов. Индустриальные центры имели преимущественное и привилегированное снабжение. Для сельского снабжения определяющим были целевые поставки в районы разработки полезных ископаемых, разработки леса, добычи пушнины и др. В распределении продуктов в определенной мере учитывался характер сельскохозяйственной специализации региона, но фактор целевых поставок для сельского снабжения был определяющим.

Виталий ПОДГАЕЦКИЙ, Юрий СВЯТЕЦ
Днепропетровский государственный университет, Украина

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ МНОГОМЕРНОГО СТАТИСТИЧЕСКОГО АНАЛИЗА В ЗАДАЧАХ ИЗУЧЕНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЙ ИСТОРИИ УКРАИНЫ ПЕРИОДА НЭПА

1. Изучение рынка и связанных с ним изменений социально-экономических отношений в аграрном секторе Украины в период НЭПа возможно на целом комплексе массовых

статистических источников: динамических переписей, бюджетных обследований, подворных выборочных опросов и др. С точки зрения методики изучение социально-экономических процессов в крестьянстве, безусловно, имеет свои корни в земской статистике, которая наиболее часто прибегала к составлению и использованию для этих задач комбинационных таблиц статистических материалов.

Последние достижения в области математической статистики, где шел активный поиск и разработка методов многомерного анализа и компьютеризации, открыли новый раздел в науке - многомерный статистический анализ, который сегодня находит свое применение и в работах историков. Примеры использования МСА в исторических исследованиях позволяют сделать вывод о возможности решения следующих задач по аграрной истории Украины в период НЭПа:

а) построение географической типологии (различные естественно-исторические исходные условия приводят к различным по уровню развития социально-экономическим отношениям);

б) построение классификации крестьянских хозяйств Украины на основе определенного математического критерия.

2. Сходные типологические задачи могут быть поставлены и для индустриально сектора украинской экономики указанного периода, в частности, представляет интерес построение различных типов социально-профессиональной структуры городского населения и выявление тенденций социально-экономического развития городов Украины.

С методической точки зрения здесь интересны такие вопросы, как:

1) возможность перепроверки результатов одного метода с помощью другого на одних и тех же исходных данных;

2) выявление ведущих признаков, определяющих основные различия социально-экономических типов городского населения и тенденций развития городов Украины в период НЭПа.

Такую возможность предоставляет исследователю метод главных компонент. Во-первых, можно установить определенную иерархию компонент в зависимости от их вклада в дисперсию. Во-вторых, анализируя положительные и отрицательные взаимосвязи компоненты с признаками, характеризующими распределение самодеятельного населения городов по социальному составу и сферам приложения труда, возможно определить ее содержательный смысл.

Наконец, располагая объекты в поле двух компонент, мы проводим своеобразную классификацию объектов, которую можно сравнивать с типологией, полученной для данного набора объектов с помощью, например, кластерного анализа.

Георгий САТАРОВ

Центр ИНДЕМ, Москва, Россия

ЛАТЕНТНЫЕ ФАКТОРЫ ПОЛИТИЧЕСКОГО РАЗМЕЖЕВАНИЯ НА СЪЕЗДАХ НАРОДНЫХ ДЕПУТАТОВ

С середины семидесятых годов в серии работ, посвященных анализу политической структуры Конгресса США, С.Б.Станкевичем и автором развивался новый подход к анализу скрытых факторов, определяющих политическое размежевание. В настоящее время Центром прикладных политических исследований ИНДЕМ этот подход совершенствуется, реализуется в виде компьютерных программ и используется при решении аналогичных исследовательских задач на материале съездов народных депутатов, а также практических задач аналитического обслуживания съездов.

Представляемый доклад посвящен описанию данного подхода, программных средств, а также основных результатов, полученных с их помощью.

Упомянутый подход состоит в построении геометрического представления политических позиций законодателей, проявленных в результатах поименных голосований; это представление основано на многомерной компенсаторной модели. В результате каждому из парламентариев ставится в соответствие точка конечномерного евклидова пространства. Размерность пространства соответствует числу латентных факторов, необходимых для описания разнообразия политических позиций. Каждая ось образует искомым латентный фактор, а координаты точек на оси - суть значения, приписанные законодателям по этому фактору. Отображение строится таким образом, что законодателям, голосующим сходным образом, приписываются близко расположенные точки, а политическим антиподам - точки, расположенные далеко друг от друга.

Процедура построения соответствия содержит следующие этапы:

- 1) статистическая оценка числа латентных факторов;
- 2) построение "грубой структуры" с помощью процедуры многомерного шкалирования, интерпретация полученного решения;
- 3) построение моделей "идеальных законодателей" и формального описания голосований в терминах выявленных факторов;
- 4) построение окончательного представления как нормированного взвешенного расстояния до "идеальных" моделей в пространстве голосований.

Реализация указанных процедур обеспечена необходимыми программными средствами, а их результат используется в двух информационно-аналитических системах. Одна из них позволяет, помимо решения традиционных поисковых задач, изучать эволюцию политических позиций отдельных депутатов и всевозможных депутатских групп. В другой системе описанное выше представление используется как основа для прогнозирования результатов выборов на отдельные посты в законодательном органе (например - выборов Председателя Верховного Совета), а также для решения других аналитических задач. Эти системы успешно апробированы на материалах съездах народных депутатов СССР и РСФСР.

Среди многочисленных результатов, полученных с помощью описанной методики и программных средств, упомянем следующие. Для съездов народных депутатов СССР со второго по пятый получены двумерные представления. Везде в качестве основного выступает фактор "консерватизм/реформизм", порожденный в основном противостоянием между "коммунистическим фундаментализмом" и демократическим радикализмом. Второй фактор определяется конкретным сюжетом съезда. В соответствии с общими закономерностями, свойственными законодательным органам, консервативное крыло было существенно сплоченнее, чем реформистское.

Российские съезды изучались, также начиная со второго. Здесь, до четвертого съезда включительно, политическое размежевание было одномерным. Единственный фактор совпал с основным фактором размежевания на союзных съездах - "консерватизм/реформизм". Однако на этих съездах поляризация была много острее, и, что самое интересное, реформистское крыло было сплоченнее консервативного. Начиная с летнего раунда пятого съезда политическая структура стала двумерной, фактор "консерватизм/реформизм" сменился факторами, коррелирующими с ним, но имеющими более ясную политическую природу.

Помимо перечисленных выше, были получены многочисленные результаты, связанные с описанием политических позиций депутатских фракций и других групп депутатов, и многие другие.

КОМПЬЮТЕРЫ В ОБУЧЕНИИ ИСТОРИИ И ПРЕПОДАВАНИИ КУРСА ИСТОРИЧЕСКОЙ ИНФОРМАТИКИ

Елена БАЛЫКИНА

Белорусский государственный университет, Минск, Белоруссия

ИСТОРИЧЕСКАЯ АЗБУКА В КОМПЬЮТЕРНЫХ ИГРАХ-ГОЛОВОЛОМКАХ

Новые информационные технологии открывают большие возможности для введения компьютеров в систему преподавания истории.

Компьютер со специально разработанными информационными и обучающими программами может с успехом взять на себя работу по "натаскиванию" студентов в знании фактического материала, например, в запоминании дат, имен, событий, т.е. определенного минимума исторических фактов, а также может оказать своим пользователям эффективную помощь в усвоении теоретических положений, т.е. понятийного аппарата.

Усвоение учебного материала будет успешным лишь при наличии семиотического базиса (знаково-понятийного материала) для построения логического каркаса знаний. С этой точки зрения важно не только усвоение "исторической азбуки", но и расширение объема словесных символов в языковых анналах обучаемых.

Решение проблемы логического установления связей между явлениями и расширения кругозора обучаемых привело к выводу, что для этой цели могут быть использованы интеллектуальные игры-головоломки (чайнворд, кроссворд, кроссчайнворд, шарада, метаграмма, анаграмма, палиндром, логогриф, ребус, криптограмма). Достоинство этих дидактических средств заключается в том, что они вносят в познавательный процесс игровой элемент, активизируют умственную деятельность, стимулируют сознательный поиск в изучаемой области. С другой стороны, их можно рассматривать как умственную гимнастику, тренировочное средство для развития мышления

Компьютер - наиболее подходящее средство для реализации интеллектуальных игр-задач, однако использование компьютерных игр в учебных целях не дает желаемого успеха (имеются в виду игры развлекающего, а не обучающего характера). Эффективно применение специальных обучающих игр.

Как показывает опыт, компьютерные игры-головоломки вызывают у обучаемых повышенный интерес и стремление достигнуть наивысших результатов, способствуют быстрому и прочному усвоению ими фактического материала, так необходимого для понимания и конструирования новых исторических концепций.

Важную роль в этой технологии обучения играет методика постановки и решения учебных задач. Игра организуется таким образом, что обучаемый вначале набирает темп, отвечает на все известные ему вопросы, а затем предельно мобилизует свои способности, получая помощь трех уровней в поисках ответа на те вопросы, по которым допустил ошибки или на которые вообще не ответил. Программа наводит его на правильную мысль либо по внешним признакам заданного объекта, либо по его качественным характеристикам, нередко напоминающим и дополняющим сведения, имеющиеся в учебных пособиях, либо путем угадывания от 1 до N букв в требуемом слове.

Познавательная деятельность стимулируется также самим режимом занятий (тренаж или контроль), а в каждом из них - оценками ответов и учетом времени, затраченного на работу, характером приза, выдаваемого за решение задачи-головоломки.

Существенное значение при разработке обучающих игр имеют:

- 1) оптимальное сочетание традиционных правил составления игр-головоломок с достоинствами и ограничениями компьютера;
- 2) выбор материала для программирования - лучше всего фактологического или понятийного характера (это обусловлено и современным состоянием исторической науки и исторического образования);
- 3) жесткая

селекция вопросов применительно к определенной аудитории; 4) разработка вариантов одной и той же программы с целью создания у каждого из обучаемых иллюзии неповторяемости "его" программы (это достигается путем изменения вида, формы, фоновых и сюжетных элементов задач-головоломок, а следовательно, и части вопросов, однако с таким расчетом, чтобы общее их количество в полной мере охватывало важнейшие события и явления компьютеризируемой темы).

Серия задач-головоломок разрабатывается по истории мировой и отечественной культуры и реализована на TURBO-PASCAL версии 5.5 для ЭВМ типа IBM PC XT/AT и ЕС-1841.

Дмитрий ИЗМЕСТЬЕВ

Московский государственный университет им.Ломоносова, Россия

СИСТЕМА РАЗРАБОТКИ ОБУЧАЮЩИХ ПРОГРАММ ПО ИСТОРИИ НА ОСНОВЕ ЯЗЫКА SMILE (Story Making Interactive Language)

Язык программирования SMILE был разработан специально для создания обучающих систем по истории. Наиболее подходящим для этой цели его делают несколько особенностей: он достаточно богат и выразителен, чтобы обеспечить историку или преподавателю возможность работы с текстами, графикой, создания диалога с компьютером; в то же время он чрезвычайно прост и позволяет создавать полноценные программы без привлечения профессионального программиста. Уже через пару часов первоначального знакомства с языком SMILE непрофессионал в программировании может получить свою первую программу с вполне современным интерфейсом: окнами, меню, графикой и т.п. Кроме того SMILE позволяет работать на любом национальном языке (русском, украинском и т.п.), заменив английские наименования всех функций языка SMILE их эквивалентами в другом языке (допускается даже перевод одного слова несколькими словами с пробелами и знаками препинания - так, как это удобно и понятно пользователю).

"17 мгновений русской истории" - пример обучающей системы по истории, подготовленной с использованием технологии SMILE. Она состоит из 17 тем, охватывающих ключевые моменты русской истории с XI по XX вв. Каждая тема включает краткий обзорный курс истории соответствующего периода, компьютерный тест, позволяющий проверить знания учащегося, а также ситуационный вопросник, позволяющий учащемуся почувствовать себя в прошлом, увидеть и ощутить окружение и обстановку изучаемого исторического периода.

Технология подготовки курса чрезвычайно проста и ясно видна из демонстрационного примера. Вы можете, опираясь на готовые образцы, легко разработать собственный курс.

Марина КОРОЛЕВА, Евгений ЗЛОБИН

Общественный Научный Центр Гуманитарных Проблем

Московский государственный университет им.Ломоносова, Россия

АВТОМАТИЗИРОВАННАЯ ОБУЧАЮЩАЯ СИСТЕМА ПО ИСТОРИИ РОССИИ

Автоматизированная обучающая система "HISTORY" предназначена для компьютерной поддержки курса истории России, читаемого на гуманитарных факультетах университетов, пединститутов, гимназий, лицеев и школ с гуманитарным уклоном.

Система включает 4 пакета программ. Тематика пакета "HISTORY-1":

1. Первобытнообщинный строй на территории страны.
2. Восточные славяне в древности. Киевская Русь.
3. Княжества в период феодальной раздробленности.

4. Образование централизованного государства.
5. Россия в середине и 2-ой половине XVII века.
6. Социально-политический кризис начала XVII века.
7. Россия в XVII веке.
8. Русская культура с древнейших времен до XVII века.
9. Россия в период петровских преобразований.
10. Российская империя во 2-ой четверти и 2-ой половине XVIII века.
11. Русская культура XVIII века. Формирование русской нации.

В пакет "HISTORY-2" входят следующие темы:

1. Россия в первой четверти XIX века. Война 1812 года.
2. Декабристы. Николаевская эпоха в истории России.
3. Социально-экономическое развитие России во 2-ой половине XIX века.
4. Великие реформы 1860-1870-х годов.
5. Демократы, революционеры и народники во 2-ой половине XIX века.
6. Россия на рубеже веков. Русско-японская война.
7. Революция 1905-1907 годов.
8. Россия между двух революций.
9. Первая мировая война. Февральская революция.
10. Русская культура в первой половине XIX века.

В настоящее время заканчивается работа над пакетами "HISTORY-3" и "HISTORY-4", которые охватывают период XX века.

Диалог "студент-компьютер" в обучающей системе строится на основе моделирования диалога "студент-преподаватель".

Каждая тема включает 50-60 контрольных вопросов. Форма ответа - свободная или выбор из предложенных вариантов. Ответ обучаемого анализируется на предмет наличия в нем ключевых слов или их синонимов. Учитывается полнота ответа. Проводится диагностика ошибочного ответа. В систему заложены типовые ошибки обучаемых и соответствующие им комментарии.

В зависимости от ответа система предлагает обучаемому дополнительные, более простые вопросы для устранения его ошибки. При ответе на вопрос можно воспользоваться справочной информацией.

Работу с системой можно прервать на любом шаге, а на следующем занятии продолжить с этого же места. В конце работы на экран выдается протокол, включающий оценки за отработанные вопросы, а для преподавателя - общий журнал контроля, в котором для каждого обучающегося записывается его фамилия, тема, количество вопросов и итоговая оценка.

Использование автоматизированных обучающих систем (АОС) наиболее целесообразно для самостоятельной работы студентов. Это обусловлено определенными их преимуществами перед традиционными средствами обучения: наличием действенной обратной связи, комфортностью обучения, учетом индивидуальных особенностей учащихся (уровень знаний, скорость восприятия информации, быстрота реакции).

Практика использования АОС "HISTORY" показала ее достаточно высокую эффективность, особенно при изучении конкретного фактологического материала: даты, имена, события и т.п. Дальнейшие направления разработки АОС по истории ведутся по линии создания экспертных систем, введения в учебный диалог средств multi-media и гипертекста.

Система реализована для ЭВМ типа IBM и СМ. Примерный объем одного пакета программ - 2 МВ.

Геральд ШПРЕНГАГЕЛЬ

Зальцбургский университет, Австрия

МЕЖДУНАРОДНЫЙ КУРС
"НОВЫЕ МЕТОДЫ В ИСТОРИИ" В ЗАЛЬЦБУРГСКОМ УНИВЕРСИТЕТЕ -
РАЗВИТИЕ ДОПОЛНИТЕЛЬНЫХ УЧЕБНЫХ ПРОГРАММ И ПРОГРАММ ДЛЯ
АСПИРАНТОВ

1. Немного истории.

Международный университетский курс "Новые методы в истории", читаемый в университете Зальцбурга проф. Г.Шмидтом, имеет целью преподавание методики и техники исследования в таких областях как:

- 1) компьютерные технологии в истории и формальные методы анализа (квантитативная история);
- 2) устная история и качественные методы в социальных исследованиях;
- 3) анализ визуальных источников, семиология и *imagology*.

Остановимся, в основном, на квантитативной секции этого курса, которая является ядром всей программы, а также на ее главных методических и дидактических принципах.

Концептуальное оформление этого курса относится к 1978 г., когда в Университете г.Линца (Верхняя Австрия) был впервые проведен не просто семинар, а "квантитативный курс" (QUANT COURSE). Университет в Линце был только что открыт, и у молодых историков были хорошие возможности попробовать свои силы в области применения количественных методов в истории. В первые годы курс QUANT поддерживался лишь небольшим кругом специалистов из Германии и Австрии, но именно в то время были сформированы его основные принципы, предполагающие:

- введение в квантитативную историю, статистические и компьютерные методы в исторических исследованиях с помощью моделирования и даже выполнения небольших, но реальных исследований;
- семинар, как форму интенсивного обучения в небольших группах в течение полной недели (а позднее - 10 дней);
- практические рекомендации по проведению исследований в области квантитативной истории. Кроме того, с самого начала курс стал междисциплинарным и проводился совместно с социологами, экономистами, политологами, психологами, филологами, демографами и т.д.

В 1981 г. курс QUANT "переехал" в Зальцбург вместе со своим создателем и руководителем Герхардом Ботцем, который получил место профессора современной Австрийской истории в университете Зальцбурга. В последующие годы схема преподавания курса изменялась и развивалась. Определелись два уровня обучения - для начинающих и "продвинутых" исследователей, появились новые разделы, посвященные качественным методам и устной истории (курс QUAL, с 1985 г.), а затем - "визуальной" истории.

В 1988 г. после десяти лет существования курса QUANT и трех лет курса QUAL была опубликована хрестоматия "Quality and quantity", базирующаяся на результатах работы множества исследователей, развивающих эту область. Схема курса к этому времени была уже хорошо отработана, и трудности первых лет были преодолены. Отношение преподавателей и слушателей к самому курсу держалось уже не на энтузиазме, а, по выражению составителей, на "спокойном профессионализме".

2. Методологические принципы.

В последние десятилетия западноевропейская историография характеризуется двумя главными тенденциями. Первая опирается на количественный подход к проблемам, основанный на применении статистических процедур и таксономических теорий социальной организации и развития; вторая - на качественный подход, который фокусирует внимание на индивидуальностях и значении отдельных явлений и событий.

Сейчас наблюдаются признаки консенсуса между этими двумя направлениями, пришедшего на смену недавним дискуссиям. Положение дел характеризуется, словами П.Бурка, "наличием ободряющих примет сближения, если не полного синтеза" этих направлений.

Например, раньше преподаватели и слушатели курса обязательно сталкивались с проблемами, осложняющими заимствование методов других дисциплин в исторических исследованиях: исторические источники с их многомерностью и сложной структурой нельзя втиснуть в стандарты кодирования и категоризации, широко применяемые в естественных и социальных науках. Успехи последних лет в развитии программного обеспечения (в частности, система KLEIO М.Таллера) позволяют представить материал как машиночитаемый источник, информация которого может быть востребована разными путями: в целях источниковедческой критики, для текстуального или семиотического анализа, а также может быть преобразована для статистической обработки. Таким образом, частью учебных программ стали методы анализа нечисловой информации и методы ее хранения, ориентированные на базы данных. Многомерный анализ данных, по-видимому, открывает новые возможности эвристического подхода. В рамках курса QUANT развиваются аналитические методы анализа текстов, а с другой стороны, в курсе QUAL работа с нарративными источниками не сводится только к специфическим методам источниковедческой критики, но и выходит на компьютерные методики анализа таких источников.

В этом как раз и может быть заключен тот синтез, о котором говорит П.Бурк. Возможно, что оживленные дискуссии о преимуществах и ценности качественных и количественных методов, ведущиеся среди преподавателей и студентов разных классов и секций Зальцбургской школы, позволят продвинуться в этом направлении.

3. Международное сотрудничество и структура университетского курса "Новые методы в истории"

С 1990 года наш университетский курс стал частью европейской сети летних школ в области преподавания новых методов в истории, а также частью европейской межуниверситетской программы ERASMUS. Преподаватели и студенты в основном набираются из таких западно-европейских стран, как Германия, Великобритания, Италия, Швейцария и др. Растущий интерес к нашему курсу среди восточно-европейских государств выдвигает университет Зальцбурга на видное место в контактах между Востоком и Западом.

В заключение доклада будут представлены возможности посещения нашего университетского курса и охарактеризованы те гранты, которые обеспечивают проживание слушателей в Зальцбурге. Кроме этого, будут представлены и распространены информационные листы для возможных участников школы следующего, 1993 года.

КОМПЬЮТЕРНЫЙ АНАЛИЗ ИЗОБРАЗИТЕЛЬНЫХ ИСТОЧНИКОВ

Владимир МОРОЗОВ, Александр ПРУЖИНИН
Лаборатория оптической телеметрии, Москва, Россия

СРЕДСТВА ВВОДА ИЗОБРАЖЕНИЙ

Хранение, поиск и анализ визуальной информации ("Image Processing") являются новым и перспективным направлением исторической информатики. Работа на компьютере с изобразительными источниками открывает широкие перспективы перед специалистами в областях искусствоведения, музееведения и других сферах, бывших традиционно отстраненными от современных информационных технологий. В отличие от текстовых источников, для которых практически не существует проблем ввода информации в память компьютера, изобразительные источники представляют значительные технические трудности в этом отношении.

Первым этапом процесса обработки изображений является ввод и перевод визуального образа в цифровую форму. Современное оборудование для решения этих задач можно разделить на две основные группы. К первой относятся сканеры, ко второй - источники видеосигнала, такие, как камеры, видеоматрицы и т.п.

Сканеры позволяют получать компьютерные копии изображений. Основным источником качественного цветного изображения являются слайды. Лаборатория оптической телеметрии имеет автоматический комплекс для обработки фотоизображений, построенный на основе микроденситометра, связанного с персональной ЭВМ. Этот комплекс обеспечивает измерение эффективной оптической плотности отдельных элементов фотоизображения и двумерного распределения прозрачности по изображению. Размеры образца могут достигать 250 на 250 мм. Диапазон регистрируемых значений оптической плотности зависит от режима работы устройства. Для черно-белого режима регистрации - от 0.02 до 4.0 D, а для цветного от 0.03 до 2.5 D. При этом обеспечивается точность порядка 0.02 - 0.07 D. Комплекс обладает очень высоким пространственным разрешением - до 5 мкм и скоростью считывания информации около 20 тысяч точек в секунду.

Программное обеспечение предназначено для управления устройством и его тестирования, автоматического сканирования по заданному алгоритму, фотометрирования и сохранения данных. Результаты представляются в виде матрицы эффективной оптической плотности или графических файлов стандартных форматов (TIFF, PCX). Максимальное количество оттенков для серых изображений - 256, для цветных - 16,7 миллионов.

Технические характеристики комплекса позволяют получать компьютерные изображения с фотоисточника без потери качества. Примером получаемых с помощью данной аппаратуры картинок может служить набор изображений икон Белоруссии XV - XVII веков. Этот набор можно просматривать на любой IBM-совместимой персональной ЭВМ с супер-VGA видеоадаптером и аналоговым VGA монитором. Изображения были получены из файла с 24-битовыми цветами аппроксимацией на палитру из 256 цветов (из 16,7 миллиона возможных). Пространственное разрешение при вводе составляло около одной десятой от максимального (50 мкм). Коллекция слайдов, послужившая источником для этого набора, была предоставлена группой энтузиастов из Академии наук Белоруссии. Эта коллекция насчитывает около 10 тысяч изображений по древней культуре Белоруссии.

Вторая группа устройств для ввода изображения состоит из различной видеоаппаратуры. Оцифровка изображения производится специальными платами, вставляемыми в слот компьютера, обычно это фреймграбберы или графические адаптеры высокого класса (TARGA, ATVista и другие). Вывод обычно быстрее сканирования и не требует

наличия слайда, однако пространственное разрешение видеосигнала ограничено по крайней мере телевизионными стандартами (около 600 линий на кадр). Эта величина приемлема для многих задач, таких, как подготовка простых демонстраций, создание каталогов. Но для качественного ввода изображений сканеры остаются пока единственным средством. Недостатком большинства современных сканеров является необходимость использования плоского носителя изображения (слайда, фотографии и т.д.) определенного размера. Решением этой проблемы могут стать сканирующие камеры, в которых вместо фотопластинки используется сканирующее устройство. Например, проекционный сканер корпорации Extel обеспечивает регистрацию 2 тысяч точек на меньший размер изображения при 256 градациях серого. С его помощью можно получать изображения любых источников, от микроорганизмов до огромных зданий.

Юрий ФИКФАК

Институт Словенской этнологии, Любляна, Словения

СЛОВЕНСКАЯ ШТИРИЯ НАКАНУНЕ МАРТОВСКОЙ РЕВОЛЮЦИИ 1848 года (Междисциплинарный проект)

В первой половине XIX в. в Словении появились два монументальных произведения, два больших сборника, которые позволяют воссоздать объемную этнографическую, историческую и географическую картину Словении того времени и, в особенности, штирийской ее части.

Во-первых, огромный интерес вызывает широта и полнота источника, находящегося в архиве земли Штирия в Граце, который известен, как серия Гетше. Словенская часть издания находится в ксерокопированном виде в Институте Словенской этнологии Словенской академии наук и искусства в Любляне. Вторым источником является Францисканский кадастр, содержащий сведения о географическом положении, имущественных отношениях, землепользовании и т.д.; благодаря точности информации этот кадастр используется в практически неизменной форме до настоящего времени. Благодаря усилиям эрцгерцога Иоганна и его секретаря доктора Георга Гета в 1840-41 гг. были опубликованы (на немецком языке) результаты семи анкетных опросов различных должностных лиц. Спустя 150 лет доктором Нико Куретом была подготовлена словенская редакция издания. Однако до настоящего времени изданы только 4 сборника с информацией о 130 населенных пунктах (из более чем 505 в рамках государственных границ Словении) После принятия решения о продолжении этой трудоемкой работы возник вопрос о создании новой концепции.

Многогранность и объемность материала позволяет выделить два уровня содержательной концепции проекта и обозначить фазы его разработки:

а) уровень общего описания (серия Гета), для которого важнейшими задачами являются продолжение словенской компиляции и критический анализ немецкой редакции;
б) уровень отдельных населенных пунктов (два населенных пункта в год), который обеспечит монографические микроисследования по каждой конкретной местности, а также откроет возможности гипертекстовой методики работы с источниками.

Для первой фазы проекта (от 120 до 130 населенных пунктов) потребуется около трех лет.

Основными методическими принципами проекта являются:

а) Принцип так называемой "верности источнику", т.е. использование всех типов источников, относящихся к рассматриваемому времени, а именно вопросов и ответов из серии Гета, других текстов, например, школьных и пасторских хроник, записок путешественников, церковных книг, карт и схем кадастра.

б) Междисциплинарные и методические инновации: привлечение специалистов из различных областей знания. Предполагается, что использование экспертов позволит

создать на основе всех задействованных источников новое представление об изучаемом хронологическом периоде. Большую роль в реализации данного проекта играет также применение новых информационных технологий, особенно компьютерных систем хранения и анализа визуальной и картографической информации. (Речь идет об использовании программного обеспечения баз данных по изобразительным источникам, разработанного под руководством д-ра Манфреда Таллера (Геттинген) и д-ра Герхарда Яритца (Кремс) - Прим. ред.).

в) Развитие международных научных связей: источники предоставляют возможность участия в проекте немецкоязычных исследователей, которые могут дать несколько иной взгляд на картину прошлого. В проекте участвуют институты Словенской академии наук, Институт истории им. М.Планка в Геттингене, Институт бытовых реалий в Кремсе, Институт этнологии в Граце.

г) Создание однородной и относительно большой базы данных, имеющей возможность гипертекстовой обработки и интегрированной в международный архив баз данных.

д) Исследовательское обучение или обучающее исследование: проект предоставляет возможность каждому заинтересованному исследователю овладеть специальными знаниями при помощи семинаров и курсов.

Герхард ЯРИТЦ, Барбара ШУХ
Институт средневековых реалий Австрийской Академии Наук
Кремс, Австрия

ПРОЕКТ БАЗЫ ДАННЫХ "REAL" ПО ИЗОБРАЗИТЕЛЬНЫМ ИСТОЧНИКАМ

Одной из основных задач Института повседневных реалий средневековья и раннего нового времени Австрийской Академии наук является анализ изобразительных источников. Фотоархив института, насчитывающий около 20000 единиц хранения, составляет основу для исследовательской работы. Фотоматериалы созданы главным образом на основе австрийских художественных образцов периода XII - XVI вв. Картины будут описываться при помощи системы KLEIO с использованием стандартизированного словаря естественного языка. Это описание, лежащее в основе модели базы данных, позволит устанавливать связи и иерархию не только формальных атрибутов полотен, но и их сюжетов.

Документирование такого рода, а также основанные на нем принципы анализа развиваются более десяти лет. Создаваемая база открыта каждому заинтересованному лицу на основе свободного доступа.

В ближайшие два года планируется работа по взаимному связыванию информации картин, находящихся в базе данных, и апробирование методов и возможностей цифровой обработки полотен в рамках исследовательской деятельности историков. Часть этих опытов основывается прежде всего на сотрудничестве австрийского Института повседневных средневековых реалий в Кремсе и Института исторических наук в Граце с Институтом истории им. Макса Планка в Геттингене. При этом немецкие партнеры прежде всего являются разработчиками программного обеспечения (KLEIO), а австрийская сторона готовит методический и содержательный научно-исторический инструментарий. Существенную поддержку при реализации проекта оказывают представительства фирмы IBM в Австрии и Англии, предоставившие свой совместный программный продукт по обработке художественных образов IMAGE ASSISTANT. Уже сегодняшний уровень развития определяет широкую сферу возможностей цифровой обработки картин, принимая во внимание тесную интеграцию изобразительных источников и исторического исследования. (см. рис. 1 и 2, полученные с экрана компьютера и дающие представление о возможностях работы исследователя одновременно с изобразительной и вербальной информацией).

Дальнейшее развитие программного обеспечения включает в себя следующие компоненты:

- 1) Immediate Image Retrieval - возможность использования картин наряду с их вербальными описаниями при помощи дружественного интерфейса в базе данных.
- 2) Image Enhancement - использование компьютера в "улучшении" полотен путем увеличения, через применение фильтров и трансформации для контурного усиления, а также для восстановления "читабельности" частично разрушенной документации.
- 3) Создание предпосылок для использования в будущем компьютерного сравнения образов и их распознавания. Это призвано инициировать методические запросы историка по включению анализа образа в канву исследования.
- 4) Связывание текстовой и изобразительной информации, полученной на основе анализа картин, позволяет, с одной стороны, использовать определенный фрагмент художественного полотна при запросе соответствующего вербального описания. С другой стороны, возникает возможность перевода фрагмента непосредственно в "архив", что облегчает интерпретацию описания его содержания.
- 5) В непосредственной связи с вышеизложенным находится применение макросов, которые устанавливают необходимую последовательность описания и обработки изображений.
- 6) В свою очередь, при реализации проекта большое значение имеет возможность интеграции с другими базами данных по иным типам источников (письменным, археологическим и т.д.), что позволяет использовать дополнительную информацию при анализе художественных полотен.

Герхард ЯРИТЦ, Барбара ШУХ
Институт средневековых реалий Австрийской Академии Наук
Кремс, Австрия