

АССОЦИАЦИЯ "ИСТОРИЯ И КОМПЬЮТЕР"
МОСКОВСКОЕ ГОРОДСКОЕ ОБЪЕДИНЕНИЕ АРХИВОВ

КРУГ ИДЕЙ: НОВОЕ В ИСТОРИЧЕСКОЙ ИНФОРМАТИКЕ

ТРУДЫ I КОНФЕРЕНЦИИ АССОЦИАЦИИ "ИСТОРИЯ И КОМПЬЮТЕР"

Издательство Московского городского объединения архивов

Москва, 1994.

ББК 73.

К84.

Редакционный Совет: д.и.н. Ю.П.Бокарев (Москва), д.и.н. Л.И.Бородкин (Москва), к.и.н. В.Н.Владимиров (Барнаул), к.т.н. З.О.Джалишвили (Санкт-Петербург), С.М.Завьялов (Москва), к.и.н. Е.В.Злобин (Москва), д.и.н. С.Г.Кашенко (Санкт-Петербург), к.и.н. Т.Л.Моисеенко (Москва), д.и.н. В.В.Подгаецкий (Днепропетровск), д.и.н. И.Р.Рафи-заде (Баку), к.т.н. Г.А.Сатаров (Москва), д.и.н. В.Н.Сидорцов (Минск), д.и.н. Т.И.Славко (Екатеринбург), к.и.н. В.С.Тяжелникова (Москва).

Ответственные редакторы:
Л.И.Бородкин, В.С.Тяжелникова

Редакция:
Т.Ф.Измestьева, Л.Л.Пушкова, В.И.Тихонов, И.Ф.Юшин

Рисунок на обложке выполнен художником В.Н.Куплевацким.

Круг идей: новое в исторической информатике. Труды I конференции Ассоциации "История и компьютер", 1994.-- 170 с.

ISBN: 5 - 7228 - 0007 - 4.

Издательство Московского городского объединения архивов, 1994
Ассоциация "История и компьютер", 1994

Оглавление

Предисловие

I. НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ

Информатика, математика, история:
"персонификация" междисциплинарного процесса
Л.И.Бородкин (Москва, Россия)

Технология разработки баз данных и
источнико-ориентированная обработка данных
В.Леверманн (Геттинген, ФРГ)

Методика оценки информационного потенциала
комплексов массовых источников
В.И.Тихонов, В.С.Тяжельникова, И.Ф.Юшин (Москва, Россия)

Российские историки XVIII - нач.ХХ вв.:
проект и информационная система
Д.А.Гутнов, В.А.Перевертень (Москва, Россия)

II. АРХИВЫ МАШИНОЧИТАЕМЫХ ДАННЫХ

Архивы машиночитаемых данных и проблемы архивирования
машиночитаемых исторических источников
И.М.Гарскова (Москва, Россия)

О современном состоянии и перспективах работы
с архивом машиночитаемых документов
А.К.Соколов (Москва, Россия)

Научно-исследовательская программа "Банк данных
по региональной истории: Урал в ХХ веке"
Т.И.Славко (Екатеринбург, Россия)

Создание банка данных по генеалогии Беларуси:
проблемы и перспективы
В.Л.Носевич (Минск, Белоруссия)

III. ОБРАБОТКА ИСТОРИЧЕСКИХ ИСТОЧНИКОВ

- Экономический барометр
Ю.П.Бокарев (Москва, Россия)
- Система для работы над материалами подворных описей
В.Л.Акимов, Н.М.Арсентьев (Саранск, Россия)
- Аграрная политика государства в фокусе
общественного мнения селян в начале 1920-х гг.
(опыт работы с программным обеспечением SPSS/PC+)
Д.Х.Ибрагимова (Москва, Россия)
- Программа FUZZYCLASS:
новые возможности археологического исследования.
(Типология поселений Алтая раннего железного века)
М.Т.Абдулганеев, В.Н.Владимиров (Барнаул, Россия)
- К вопросу об авторстве предсмертного письма
Б.В.Савинкова - опыт комплексного исследования
А.В.Быстров, Е.В.Злобин (Москва, Россия)
- Об изучении социальной дифференциации
закавказского крестьянства в конце XIX в.
(По материалам бюджетного обследования)
И.Р.Рафи-заде (Баку, Азербайджан)
- Вариант создания базы данных о крестьянских хозяйствах
Украины в годы нэпа
В.В.Подгаецкий, Ю.А.Святец (Днепропетровск, Украина)

IV. КОМПЬЮТЕР В ОБРАЗОВАНИИ ИСТОРИКА

- Многомерность представления знания
в информационных технологиях обучения
Н.И.Миницкий (Минск, Белоруссия)
- Компьютерная технология обучения истории
в Белорусском государственном университете:
пути реализации
Е.Н. Балыкина, В.Н. Сидорцов (Минск, Белоруссия)

Сведения об авторах

ПРЕДИСЛОВИЕ

Информационное творчество человечества охватывает все новые и новые сферы, постепенно приобретая всеобъемлющий характер. Возникает новая информационная культура, гуманитарная по своей сути - она рождается в результате усилий людей, изменяя их и формируя новую социальную среду. В контексте такой культуры роль профессиональных историков имеет две отличительные черты: они активно участвуют в созидании информационного богатства, с одной стороны, и стремятся систематизировать и обобщить то, что накапливается в других областях научного знания и сферах человеческой жизни, с другой. В обоих случаях существенным становится охранительное начало - и при переводе в компьютер знаний о прошлом, и при обобщении вновь возникающих реалий. Именно специалисты по исторической информатике в различных странах взяли на себя нелегкую заботу о сохранении и систематизации, изучении информационного потока, который, лавинообразно нарастая, может захлестнуть современников и исчезнуть бесследно для потомков. Важным источником информации о нашем времени будут базы данных.

Масштабность обозначенных задач потребовала интеграции усилий, направленных на разработку и использование новых информационных технологий работы с историческими источниками. Уже к середине 80-х годов специалисты по исторической информатике ряда стран Западной Европы пришли к выводу о необходимости институализации международного научного сообщества, складывающегося в этом активно развивающемся направлении исследований. В 1986 г. была создана Международная ассоциация "History & Computing" (АНС), которая к 1993 г. объединила полтора десятка национальных ассоциаций, включающих около 1000 членов. Сегодня это одна из самых представительных профессиональных ассоциаций историков и, без сомнения, самая активная из них. С 1986 г. АНС проводит ежегодные международные конференции, в которых участвуют сотни специалистов из разных стран, публикует объемистые тома трудов этих конгрессов (а накануне этих крупных форумов проводит серию небольших проблемно-ориентированных рабочих семинаров), инициирует совместные исследовательские проекты и разработку специализированного "источнико-ориентированного" программного обеспечения, участвует в организации летних школ по данной проблематике, отработывает оптимальную модель курса исторической информатики и т.д. (Отметим один из результатов этой деятельности АНС: в ряде университетов Западной Европы на исторических факультетах введена специализация по профилю "History & Computing", дающая возможность получить магистерский или

даже докторский диплом по этой специальности). С 1989 г. АНС выпускает (в Оксфорде) журнал "History and Computing".

Однако в этой активной профессиональной деятельности специалисты из стран Восточной Европы практически не участвовали (достаточно сказать, что представителей нашей страны не было ни на одной из первых семи конференций АНС). Ситуация начала изменяться с января 1992 г., когда АНС организовала международный симпозиум по координации исследовательской деятельности в области исторической информатики в странах Западной и Восточной Европы, который состоялся в Зальцбурге (Австрия). Участники симпозиума, прибывшие в Зальцбург из 11 европейских стран, обсудили современное состояние исторической информатики, наметили основные направления международного сотрудничества. Работа симпозиума была конструктивной (чему во многом содействовал президент АНС д-р Манфред Таллер), проходила в атмосфере полного взаимопонимания специалистов из разных стран. Результаты этой международной встречи не замедлили сказаться. Уже к лету 1992 г. в ряде стран Восточной Европы были созданы ассоциации "History & Computing". Десятки аспирантов и молодых ученых из стран Восточной Европы (в том числе из России) смогли принять участие в Зальцбургской летней школе-92, посвященной новым методам исторических исследований. В июне того же года в Ужгороде состоялась международная конференция "Новые компьютерные технологии в исторических исследованиях и образовании", которая была организована Закарпатской ассоциацией молодых историков и Историческим факультетом МГУ (представленным лабораторией исторической информатики), при поддержке АНС. Ужгородский семинар собрал 40 специалистов по исторической информатике из 13 стран. Три дня напряженной работы семинара вместили около 30 докладов и десять демонстраций оригинального программного обеспечения. Впервые ученые Восточной и Западной Европы обменялись опытом своих разработок в области исторической информатики.

На Ужгородском семинаре возникла инициатива создания ассоциации "История и компьютер" (АИК) в нашей стране. Учредительное собрание, которое состоялось затем в Москве, выбрало Совет ассоциации, который осенью 1992 г. провел ее регистрацию. Уже в октябре 1992 г. ассоциация "История и компьютер", Исторический факультет МГУ и АНС организовали первую в России международную осеннюю школу "Историческая Информатика: Европейская Модель" (осенью 1993 г. школа была с успехом проведена второй раз и становится теперь ежегодной). АИК установила прочные связи с голландской ветвью АНС, включилась в международные исследовательские проекты в области исторической информатики, приняла активное участие в подготовке двух сборников статей, отражающих тенденции развития данного направления в странах Восточной и Западной Европы.

Создание АИК способствовало установлению прочных контактов специалистов в области исторической информатики из разных стран СНГ. К концу 1993 г. ассоциация объединяла около 100 исследователей, работающих в 24-х вузах, академических институтах, архивных учреждениях Азербайджана, Белоруссии, Киргизии, России, Украины. Первая их встреча состоялась в апреле 1993 г. на I ежегодной конференции АИК в Подмоскowie.

Сборник, который предлагается вниманию читателя, в целом отражает содержание I конференции Ассоциации "История и компьютер", хотя и дополнен некоторыми статьями, поступившими позднее.

Первый раздел посвящен новым информационным технологиям в исторических исследованиях. Он открывается статьей Л.И.Бородкина, в которой автор рассматривает один из важных методологических аспектов междисциплинарного взаимодействия, возникающего "на стыке" истории, математики и информатики; автор обращает внимание на различие характера процессов внедрения в историческую науку математических методов и методов информатики. В этой связи в работе определяются контуры профессиональной подготовки специалистов в области исторической информатики. Статья, в определенном смысле, подводит итог современному состоянию проблемы, дискуссия по которой, на наш взгляд, будет продолжена на II конференции Ассоциации.

Трепетное отношение к историческому источнику (наряду с учетом исторического внеисточникового знания) можно с полным основанием считать доминирующей тенденцией развития исторической информатики в последнее десятилетие. Именно это и объединяет все остальные статьи первого раздела. Так, статья сотрудника Института истории общества Макса Планка (Геттинген, ФРГ), участника I Конференции Ассоциации В.Леверманна посвящена вопросам технологии разработки баз данных в контексте внимания к специфике исторического источника и исторических данных, в ней подробно разбираются те требования к программному обеспечению, которые предъявляет историк, стремящийся к адекватности базы данных и первичного материала. В русле подхода, ориентированного на источник, выполнена и статья В.И.Тихонова, В.С.Тяжелниковой, И.Ф.Юшина. Здесь предлагается компьютерная модель источника, однако главный акцент делается на количественный подход к внутренней критике комплекса массовых документов. Существенным новым элементом здесь является то, что компьютерная модель и пилотажная выборка выступают как средство познания внутренней природы источника, "рисунка" его информационных компонентов, что, по мнению авторов, позволяет измерить информационный потенциал массива.

Источнико-ориентированные технологии можно рассматривать несколько шире, чем информационные системы для обработки исторических данных. На наш взгляд, это новая идеология отношения к информации вообще, ставящая в центр внимания сохранение инфраструктуры данных (структурной,

семантической), а также позволяющая воспроизвести историографическую и культурную среду исследовательского процесса. Ярким примером такой попытки является статья Д.А.Гутнова и В.А.Перевертенья, завершающая первый раздел. Заметим, что все указанные разработки выполнены в так называемых "гибких" системах, что знаменует новую тенденцию в исторической информатике и может рассматриваться как признак начавшейся специализации исторической информатики в совокупности других информационных дисциплин.

Отличительной чертой предлагаемого сборника является первое обстоятельное обращение отечественных специалистов к проблеме машиночитаемых архивов. Фундаментальное, первое в русскоязычной литературе, обобщение опыта работы с машиночитаемыми источниками и различных подходов к их систематизации представлено в работе И.М.Гарсковой. Уже сейчас можно с уверенностью говорить о том, что это заметный вклад не только в историческую информатику, но и основа для начала практической работы с машиночитаемыми документами в России и странах СНГ. Важным и интересным представляется взгляд А.К.Соколова на отечественный опыт собирания машиночитаемой информации, анализ как наметившихся в последние годы тенденций, так и того, что можно уже назвать историей машиночитаемых коллекций. В тесной связи с этой работой можно рассматривать статью Т.И.Славко, где подводится итог работы над региональным банком данных по истории Урала XX века. В качестве отличительной черты уральского подхода можно назвать источниковую доминанту, которая проявляется в особом внимании к фондам массовых источников уральских архивов. Работа белорусского автора В.Л.Носевича завершает раздел презентацией проекта по созданию национального генеалогического банка данных.

Традиционным можно считать третий раздел, посвященный обработке исторических данных. Но и здесь можно назвать некоторые отличительные черты, проявившиеся в последние годы. Среди них, на наш взгляд, главная - подход к проблемам обработки данных через информационную систему, через реализацию продвинутых возможностей таких систем и создание оригинального программного обеспечения под конкретные потребности обработки. Работы авторов, создающих программное обеспечение самостоятельно (Ю.П.Бокарева, В.Л.Акимова и Н.М.Арсентьева, А.В.Быстрова и Е.В.Злобина) представляют несомненный интерес как вариант реализации информационных запросов историков-исследователей, обусловленный не только спецификой источника как материала, но и особой природой исследовательской задачи, которую они пытаются решить. В этом же контексте можно рассматривать и статьи авторов, не создающих программное обеспечение самостоятельно, но строящих обработку данных на продвинутых возможностях уже опробованных историками систем (Д.Х.Ибрагимова, М.Т.Абдулганеев и В.Н.Владимиров). Две работы представляют традиционное количественное направление

(И.Р.Рафи-заде, В.В.Подгаецкий и Ю.А.Святец), в центре внимания которых конкретная историческая проблема, а методический аппарат выступает как инструмент ее изучения. Но и здесь следует подчеркнуть, что он становится все более сложным и чувствительным, включает авторские разработки.

Сборник заканчивается разделом, посвященным новым информационным технологиям в обучении. Все авторы – представители республики Беларусь, что закономерно, поскольку именно здесь достигнуты значительные успехи не только в практическом обучении историков с помощью авторских разработок (Е.Н. Балыкина, В.Н. Сидорцов), но и в осмыслении роли компьютера в обучении знаниям в более широком смысле (Н.И.Миницкий).

На наш взгляд, содержание предлагаемого сборника отражает наиболее существенные тенденции развития исторической информатики стран бывшего СССР, книга будет интересна всем, кому небезразличны проблемы гуманитарного знания в свете информационного подхода.

Отв. редакторы

**І. НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ**

**ИНФОРМАТИКА, МАТЕМАТИКА, ИСТОРИЯ: "ПЕРСОНИФИКАЦИЯ"
МЕЖДИСЦИПЛИНАРНОГО ПРОЦЕССА**

Л.И.Бородкин

Микрокомпьютерная революция 80-х гг. и внедрение новых информационных технологий в практику исторических исследований привели к появлению новой междисциплинарной области - исторической информатики. Развитие исследований в этой области требует наличия специалистов, обладающих знаниями на стыке истории и информатики. Какой должна быть траектория формирования таких специалистов? От истории - к информатике? Или от информатики - к истории? Или, может быть, историки и "информатики" просто должны работать вместе - в составе одной лаборатории, кафедры или группы? Для ответа на этот вопрос автор предлагает обратиться к опыту, накопленному в период 70-80-х гг. в области применения количественных методов в исторических исследованиях, рассмотрев его в более широком контексте математизации научного знания.

Рассматривая сложное и многоаспектное явление математизации современной науки /1/, методологи, как правило, подчеркивают значение одного из аспектов этого междисциплинарного процесса, связанного с конкретными путями его реализации, с его "персонификацией", если так можно выразиться. Об этом весьма выразительно пишет М.А. Розов: "Математизация породила в науке если не особую профессию, то особую роль, особую фигуру, фигуру математизатора. Это человек, работающий на стыках наук, математик, ставший биологом или гуманитарием и в то же время сохранивший установки и принципы математического мышления. Он призван как бы сидеть на двух стульях, согласуя то, что, вообще говоря, трудно согласуется; нередко это роль конфликтная, требующая большой разносторонности и этической или аксиологической культуры" /2/.

Интересующий нас аспект нашел отражение в публикациях, посвященных процессу математизации в различных областях социально-гуманитарного знания - в психологии /3/, социологии /4/ и т.д. /5/. Так, Ю.Н. Толстова, обсуждая требования к исследователю, работающему в области математической социологии, подчеркивает необходимость "сочетания глубокого проникновения в содержательную область с пониманием возможностей математических методов. Соединение в лице одного исследователя знания таких разных научных дисциплин как математика и социология, естественно, требует особой психологической организации такого исследователя. Он в определенной мере должен обладать качествами математика (уметь мыслить формальными категориями, достаточно глубоко разбираться в сути применяемых методов и т.д.) и качествами социолога (уметь видеть за наблюдаемыми данными живых людей, реальные социальные процессы, достаточно глубоко разбираться в изучаемом социологическом предмете). То же единение социологии и математики может осуществляться и другим путем - посредством налаживания тесного контакта между социологом и математиком, что также весьма непросто и предъявляет соответствующие психологические требования и к тому и к другому... Неудовлетворительное положение дел с применением математики в социологии, на наш взгляд, в значительной мере объясняется тем, что имеет место и наличие психологического барьера между социологом и математиком, и практическое отсутствие специалистов, соединяющих в своем лице качества и того, и другого" /6/.

Авторы другой работы о математизации социологических исследований, рассматривая цели применения математических методов, отмечают: "Ясно, что достижение всех сформулированных целей невозможно без самого тесного контакта социолога и математика. Мы считаем это главным методологическим принципом применения любого математического метода в социологии... Именно отсутствие достаточно тесных контактов между социологом и математиком (оба лица, конечно, могут совмещаться и в одном человеке) в соответствующих "болевых точках" приводит к неадекватности используемой формальной модели и, как следствие, к резкому снижению эффективности применения математического метода" /7/.

Что касается применения математических методов в исторических исследованиях, то проблема адекватного сочетания профессиональных знаний историка и математика начала обсуждаться более 20 лет назад. Так, эту проблему затрагивали в своих выступлениях на симпозиуме по актуальным проблемам источниковедения (Таллин, 1972 г.) В.М. Селунская: "В деле внедрения количественных, в том числе математических методов в историю необходимо достичь такого уровня обоюдной подготовки историка и математика, чтобы они могли вести свободный диалог, разрабатывая программу исследования избранного объекта, отбирая оптимальные математические приемы количественного

анализа" /8/, и И.Д. Ковальченко: "Очень важной является проблема взаимоотношений историка и математика. Здесь должна быть дифференциация функций и одновременно их интеграция. Историк может не иметь математического образования, но он должен понимать логическую основу применяемых методов, чтобы быть полезным математику для определения проблематики и направлений работы и суметь интерпретировать полученные результаты. Математик же обязан вникнуть в основную суть и содержание поставленных проблем" /9/.

Вопрос не потерял своей актуальности и сегодня. Так, К.В. Хвостова высказывает предположение, что "если бы в современном обществе имело бы широкое распространение комплексное образование, при котором математики серьезно изучали бы историю, а историки - математику, то проблема малой значимости математизации для истории не только бы оказалась снятой, но возникли бы предпосылки для выработки многих методик, улавливающих специфику исторических явлений" /10/.

Как нам представляется, развитие в нашей стране нового направления исследований, связанного с применением математических методов в исторической науке, показало пример достаточно успешного сотрудничества историков и математиков. Достижения сложившейся в данной области школы признаются и зарубежными специалистами-квантификаторами /11/. Так, Конрад Ярауш, сопредседатель Международной комиссии по применению количественных методов в исторических исследованиях (INTERQUANT), выделяет 5-6 национальных школ в области квантитативной истории, в том числе и советскую. Оценивая ее уровень (и используя при этом как позитивные, так и критические высказывания), К. Ярауш, в частности, отмечает: "Тесное сотрудничество (советских историков - Л.Б.) с математиками обеспечило высокие стандарты исследований в ряде областей (например, связанных с применением методов многомерной классификации, распознавания образов)" /12/.

Эти результаты являются во многом следствием целенаправленной линии на организацию лабораторий соответствующего профиля в академических институтах и университетах. Штат таких лабораторий включал и историков, и специалистов в области прикладной математики и кибернетики. "Это - реальное выражение развития комплексности в исторических исследованиях, - отмечает И.Д. Ковальченко. - Как показывает имеющийся опыт, такая форма сотрудничества обеспечивает наиболее эффективное применение математических методов в исторической науке, и ей принадлежит будущее. Естественным следствием органичного слияния усилий историков и математиков являются их совместные научные публикации, которые все более широко распространяются. Очевидно, что для успешной совместной работы не только историк должен обладать определенной математической подготовкой, но и математик - историческими знаниями как таковыми. Как правило, это оказывается вполне посильным для тех и других. Но кроме этого каждый из них,

оставаясь специалистом в своей области, должен овладеть новым стилем научного мышления: историк - математическим, а математик - историческим..." /13/.

Последнее замечание представляется весьма существенным. Действительно, разобраться в тонкостях сложного математического метода подчас нелегко и профессиональному математику, а уж разработка нового математического инструментария, учитывающего специфику данной предметной области, может потребовать от математика очень высокой квалификации. И - добавим - глубокого понимания сути соответствующей содержательной проблемы, что достигается при совместной работе историков и математиков /14/.

* * *

Останется ли верным все, что было сказано выше, если заменить слова "математика" и "математизатор" на "информатику" и "информатизатор" соответственно? Думается, что ответ на этот вопрос не однозначен (и, более того, скорее "нет", чем "да").

Дело в том, что, используя термин "информатика", мы имеем в виду в основном прикладную информатику, ориентированную преимущественно на практическое использование компьютерных технологий. Овладение этими технологиями может потребовать от историка пытливого ума, терпения и настойчивости, но, как правило, не потребует специфического (математического) стиля мышления (хотя последний и в данной ситуации будет не лишним). Овладев навыками работы с одной системой управления базами данных, историк без особого труда переходит к другой СУБД; разобравшись с одним пакетом статистических программ, можно быстро научиться работе с другим и т.д. Конечно, многое здесь зависит от индивидуальных способностей (как, впрочем, и в любом другом деле), от уровня сложности программного обеспечения, от готовности преодолеть психологический барьер и т.д., но это проблемы другого характера, чем отсутствие необходимого "стиля мышления".

Здесь, однако, может возникнуть вопрос: а как быть с разработкой специализированного программного обеспечения, учитывающего специфику предметной области (в данном случае - исторических исследований)? Справится ли историк с такой задачей? Ответим уверенно - да. Многим читателям, наверное, известны названия систем KLEIO, SOCRATES, WINCENS (этот ряд можно продолжить); указанные пакеты относятся к источно-ориентированному программному обеспечению /15/. Они были созданы на различных языках программирования историками, не имеющими специального образования в области информатики или математики.

Эти примеры не означают, разумеется, что математикам-компьютерщикам в исторической информатике "нечего делать". Напротив, можно считать удачей то обстоятельство, что десяток специалистов этого

профиля связан с отечественной исторической информатикой (пожалуй, это единственный пример такого рода на фоне ситуации в других странах). Это особенно важно в силу того, что исторические факультеты наших университетов по-прежнему очень медленно обзаводятся компьютерами (и в этом не их вина), и наличие двух-трех лабораторий, хорошо укомплектованных квалифицированными специалистами и компьютерами продвинутых моделей, создает возможности для быстрого освоения новых информационных технологий.

Не трудно предвидеть, однако, что уже в ближайшие годы проблема обеспечения вузов страны компьютерами будет в той или иной мере решена, и тогда на первый план выйдут вопросы обучения студентов-историков новым информационным технологиям. Исходя из имеющегося опыта (как отечественного, так и зарубежного), представляется необходимым предусмотреть 2-3 уровня в структуре данной образовательной компоненты, с тем чтобы верхний уровень подготовки (например в рамках магистерской программы) давал бы выпускнику-историку квалификацию специалиста в области исторической информатики, характеризуя которого Манфред Таллер пишет: "Совокупность задач, за разрешение которых должны взяться историки будущего, заставляет нас рассмотреть роль человека - специалиста, который помогает историкам использовать программное обеспечение для решения их задач. Роль человека, который применяет методы формального анализа к материалу источника выявляя структуры, специфические для исторических источников и делая возможным описание общих решений задач их обработки на современных компьютерах. И наконец, роль человека, который создает программы для решения тех проблем исторического исследования, где не помогают стандартные подходы, ориентированные на более ясную и недвусмысленную информацию, чем та, что встречается в туманном королевстве исторического знания " /16/.

В целом ряде европейских университетов в течение последних лет формируются различные модели подготовки специалистов по применению новых информационных технологий в исторических исследованиях; координация этих усилий осуществляется международной ассоциацией АНС, организовавшей в 1992-93 гг. два международных семинара по обсуждению образовательных программ по профилю "Historical Computing". Материалы этих семинаров дают представление о структуре и содержании дисциплин в рамках указанного цикла, реализованного в 16 университетах девяти стран Западной и Восточной Европы /17/, [18]. Особый интерес представляют программы подготовки магистров по данной специальности в университетах Лондона и Глазго (годовые программы, завершающиеся защитой магистерской диссертации на степень MA в первом случае и MPhil - во втором).

Очевидно, однако, что дальнейшее расширение и углубление образовательных программ по исторической информатике связано с

наращиванием исследовательского потенциала в этой развивающейся междисциплинарной области, разработкой новых методов и информационных технологий работы с историческими источниками.

ПРИМЕЧАНИЯ.

1. Так, М.А. Розов отмечает в этой связи: "Очевидны, например, социальные и социально-психологические последствия математизации. Она приводит к перестройке организационной структуры науки, меняет систему образования, разрушает иногда вековую обособленность отдельных дисциплин...". См.: Розов М.А. Философские аспекты проблем математизации науки // Математизация современной науки: предпосылки, проблемы, перспективы. М., 1986. с. 14.

2. Там же. с.14.

3. Журавлев Г.Е. Системные проблемы развития математической психологии. М., Наука, 1983; Паповян С.С. Математические методы в социальной психологии. М., Наука, 1983; Математическая психология: методология, теория, модели. М., Наука, 1985.

4. Толстова Ю.Н. Логика математического анализа социологических данных. М., Наука, 1991; Андреенков В.Г., Аргунова К.Д., Паниотто В.И. и др. Математические методы анализа и интерпретация социологических данных. М., Наука, 1989.

5. См., например: Моисеев Н. Н. Математика ставит эксперимент. М., Наука, 1979, с. 97, 106.

6. Толстова Ю.Н. Указ. соч., с. 17-18.

7. Андреенков В.Г., Аргунова К.Д., Паниотто В.И. и др. Указ. соч., с. 8-9.

8. См.: Материалы симпозиума по актуальным проблемам источниковедения (Таллин, 1972) // Источниковедение отечественной истории. М., Наука, 1977, с. 258.

9. Там же, с. 264.

10. Хвостова К.В. Гносеологические предпосылки современной количественной истории // Россия и США на рубеже XIX - XX столетий (Математические методы в исторических исследованиях). М., Наука, 1992, с. 8.

11. Jarusch K.H. The International Dimension of Quantitative History: Some Introductory Reflections // Social Science History, 1984, v. 8, p. 115-136; Jarusch K.H. (Inter)national Styles of Quantitative History // Historical Methods, 1985, v. 18, N. 1, p. 13-19; Gatrell P. and Lewis R. Russian and Soviet economic history // Economic History Review, XLV. 1992, N. 4, p. 743-754; Rowney D.K. Soviet Quantitative History // Soviet Quantitative History / Ed. by D.K.Rowney. Beverly Hills, 1984.

12. Jarausch K.H. (Inter)national Styles..., p. 15. В оценочном плане представляется интересной также следующая цитата из работы К. Ярауша: "Различия между подходами, развиваемыми национальными школами квантитативной истории, оказываются, вопреки ожиданиям, на удивление существенными. Можно даже говорить о соревновании между американской (historical social science), французской (Annales) и российской парадигмами, особенно в развивающихся странах, которые импортируют не только машины, но также и методы/методологию". Там же, с. 17.

13. Ковальченко И.Д. Методы исторического исследования. М., Наука, 1987, с. 324.

14. Разумеется, существуют примеры и другого рода, когда построение достаточно сложных математических моделей производится историком, глубоко изучившим соответствующий математический аппарат (об этом см., например: Бородкин Л.И. Историческая информатика в СССР/России: ретроспектива, состояние, перспективы// История и компьютер: Новые информационные технологии в исторических исследованиях и образовании. Ред. Л.Бородкин и В.Леверманн. St. Katharinen, 1993, с.258.

15. О системах KLEIO, SOCRATES и WINCENS см., например, статьи М.Таллера, Л.Брере и Я.Олдерволла в сборнике: История и компьютер: новые информационные технологии в исторических исследованиях и образовании. St. Katharinen, 1993.

16. Таллер М. Что такое "источнико-ориентированная обработка данных"; что такое "историческая информатика" // История и компьютер: новые информационные технологии в исторических исследованиях и образовании. St. Katharinen, 1993, с. 18.

17. Towards an International Curriculum for History and Computing (Eds. D.Spaeth, P.Denley, V.Davis, R.Trainor). Max-Planck-Institut fur Geschichte, Gottingen, 1992.

18. The Teaching of Historical Computing: An International Framework (Eds. V.Davis, P.Denley, D.Spaeth, R.Trainor). Max-Planck-Institut fur Geschichte, Gottingen, 1993.

ТЕХНОЛОГИЯ РАЗРАБОТКИ БАЗ ДАННЫХ
И ИСТОЧНИКО-ОРИЕНТИРОВАННАЯ ОБРАБОТКА ДАННЫХ

В. Леверманн

Эта статья посвящена применению технологии баз данных (БД) в историческом исследовании и дает краткий обзор современных тенденций и разработок, ведущихся в этой области. Однако здесь не будет идти речь о БД в исторических исследованиях в целом, поскольку БД могут быть полезны для историков по различным причинам: например, они могут служить основой для информационных систем типа "Банк данных по средневековью и раннему новому времени", которая содержит информацию о денежных единицах и мерах, ценах и заработных платах /1/; кроме того они могут использоваться для составления библиографий, каталогов и т.п. В этой статье основное внимание уделяется применению технологии БД в сфере источник-ориентированной обработки данных (ИООД). Прежде всего мы дадим краткое объяснение того, что именно подразумевается под "источник-ориентированной обработкой данных".

Применение компьютерных методов в историческом исследовании требует значительной предварительной работы по обеспечению возможности использования формальных операций - особенно, если обрабатывается текстовая информация, которая, вообще говоря, является наиболее распространенным типом исторических источников. Математическая ориентация первых пакетов программ обуславливала необходимость кодирования, стандартизации и классификации исторической информации до начала работы с компьютером. Известно, что обработка и анализ текстовых данных с помощью статистических методов предполагают, что информация должна быть приведена к количественным данным (цифрам и числам), т.е. к четко определенному и точному представлению. Хотя мы и не отрицаем ни пользы категоризации и классификации в историческом исследовании, ни преимуществ количественного подхода к историческим проблемам вообще, процесс кодирования и категоризации информации, очевидно, является одним из центральных мест в историческом исследовательском проекте, и - я бы добавил - часто это слабое место.

Известно также, что некоторые критики количественного подхода доказывают, что историческая информация - по крайней мере в большинстве случаев - не может быть адекватно преобразована в четко определенные понятия. Но даже, если согласиться, что мы способны представить элементы исторической информации подходящими понятиями, то все же не вызывает сомнений, что ее подготовка в соответствии с требованиями некоторого формального метода (особенно статистического анализа) является одним из наиболее сложных и длительных этапов

исторического исследовательского проекта. Этот этап имеет, очевидно, огромное влияние и на способ, которым может быть выполнен анализ, и на результаты исследования.

Тот факт, что в количественных исследованиях возможность представления исторической информации точными понятиями и категориями иногда неявно предполагается, но не обсуждается открыто, не вызывает больших проблем. Хуже то, что такие действия часто недостаточно документированы, и что еще более важно - после того как историческая информация классифицирована тем или иным образом и введена в компьютер, обычно остается мало возможностей для изменения схемы классификации, переупорядочивания информации и, в особенности, для добавления информации, не отобранной для анализа в соответствии с первоначальной схемой.

Таким образом, исследовательский проект такого рода в основе своей недостаточно гибок; он как бы опирается на неподвижный фундамент, что несомненно противоречит природе процесса исторического исследования. Этот процесс итеративен и требует постоянной модификации (фальсификации или верификации) предположений и гипотез - включая те, которыми определялся отбор информации и способов преобразования отобранной информации в понятия и категории; часто уже во время работы над проектом происходит смещение акцентов в стратегии исследования.

Осознание ограниченности традиционного способа подготовки исторической информации для автоматизированной обработки - прежде всего и главным образом для статистического анализа - привело в начале 80-х годов к возникновению источник-ориентированного подхода. Основная его идея состояла в том, чтобы преодолеть вышеупомянутые ограничения, придать большую гибкость обработке исторической информации, предоставить историку более широкий набор аналитических средств, и была тесно связана с развитием новых методов и инструментария автоматической обработки текстов (полнотекстовые информационные системы, системы управления базами данных - СУБД). По М.Таллеру, одному из пионеров этого подхода, источник-ориентированная обработка данных - в противоположность так называемой метод-ориентированной обработке данных - "это стремление смоделировать в компьютере весь объем информации, содержащейся в историческом источнике; это попытка обеспечить возможность использования источника в самых разнообразных целях. Предоставляя историку средства для различных типов анализа, он не требует от него на этапе создания БД принятия решения о том, какие методы будут позднее использованы" /2/. С одной стороны, ИООД нацелена на максимально точное сохранение источника и старается избегать всевозможного кодирования, стандартизации и т.п. на стадии ввода данных. С другой стороны, предоставляется широкий диапазон средств, позволяющих использовать различные аналитические подходы к источникам. В рамках ИООД

количественный анализ рассматривается лишь как один из альтернативных подходов, при котором компьютер используется как мощный инструмент преобразования информации в требуемые для статистического анализа понятия и категории.

Имеет смысл более детально описать некоторые средства, разработанные для различных аналитических подходов, - извлечение информации (в том числе из полнотекстовых систем), генерация отчетов, связывание записей, составление тематических справочников, обработка изображений, экспорт данных в статистические пакеты и т.д. Следовательно, мы должны сконцентрировать внимание только на первом аспекте ИООД, т.е. на том, как смоделировать в компьютере всю сумму информации, содержащейся в источнике. В этом отношении использование БД всегда играло важную роль, поскольку исторические данные - это часто структурированные данные, применение к которым технологии БД выглядит весьма естественным.

Так, с чисто научной точки зрения хорошим стилем было бы представление исторического источника в компьютере с той степенью полноты и так близко к оригиналу, насколько это возможно. Но сохранение "верности оригиналу" порождает значительные проблемы, связанные с автоматической обработкой данных. В то время как коммерческие данные обычно можно легко адаптировать к требованиям СУБД, исторические данные изначально не подходят для такой обработки. Так что, если мы действительно пытаемся избежать всех видов стандартизации на этапе ввода, то мы просто должны допустить возможность того, что наши данные часто содержат особенности, которые в общепринятых БД рассматривались бы как "нерегулярности" или "несоответствия". Упомянем лишь некоторые из них: отдельные записи могут быть очень короткими - содержащими только одно слово, в то время как другие (того же типа) содержат целые фразы, т.е. однородные объекты могут значительно различаться по длине. Довольно часто запись содержит не один признак, образующий поле БД, а два или более, имеющих одинаковый логический статус (например, когда персона имеет две или более профессий). В других случаях СУБД могут администрировать большое число атрибутов, но записи часто встречаются с пропусками. Иногда в БД должна быть включена информация, которая каким-то образом связана с наиболее интересными для нас данными, но которая не может напрямую быть внесена ни в одно из существующих полей; этот случай требует какого-то способа интегрирования "ассоциированных" данных.

Перечисленные проблемы касаются лишь структурных свойств данных. Однако наиболее серьезные вопросы связаны с предполагаемым способом использования данных в историческом исследовании: обычно историческая информация требует комплексной интерпретации. Центр тяжести здесь лежит в области "полной" или "содержательной" интерпретации, поскольку очевидно, что каждый текст, написанный на естественном языке, требует

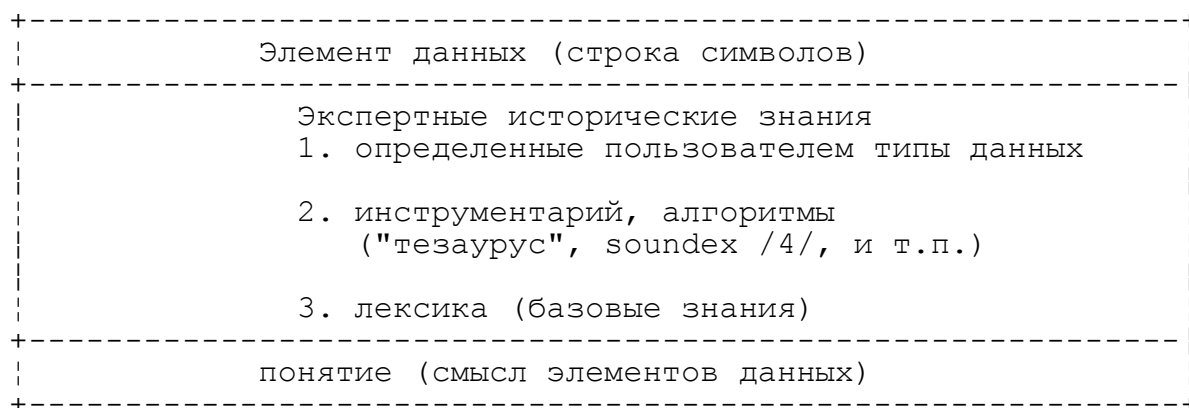
некоего рода интерпретации для понимания его смысла; даже телефонный справочник, содержащий весьма ограниченный набор данных и имеющий простую упорядоченную структуру, "интерпретируется" приложением лингвистических знаний к тексту, которые главным образом касаются букв и чисел, организованных в столбцы. То же самое, но в гораздо более сложном смысле, приложимо к интерпретации исторических источников: понимание исторического текста обычно требует привлечения разнообразной дополнительной информации - экспертного исторического знания, которое необходимо, например, для идентификации географических терминов, встречающихся в источнике, для сопоставления сумм, заданных в разных денежных единицах, для приведения дат, указанных в стилях местных епархиальных календарей, к стандартным календарным датам или для обработки различных вариантов написания имен, когда некое лицо названо в разных вариантах документов похожим (но не одинаковым) образом. Для таких случаев требуется специальное программное обеспечение, которое определено выходит за пределы традиционных СУБД.

Так, должно быть очевидным, что традиционные СУБД - разработанные для совершенно иных целей - часто не удовлетворяют потребностей обработки исторических данных. Иногда эти системы накладывают жесткие ограничения на длину полей или не обеспечивают доступа к двум или более элементам данных одного логического статуса (в прошлом мы упоминали, что реляционные базы данных, будучи наиболее популярными для коммерческих приложений, не позволяют вводить два или более элемента в поле БД). Пропуск значений также может вызвать проблему, т.к. у нас нет возможности "ассоциировать" имеющуюся информацию с содержимым поля БД. Хотя это касается "нерегулярности" или "несогласованности" исторических данных, то же самое, конечно, верно и для второго из вышеупомянутых аспектов /3/.

В самом деле, похоже, что не существует больших разногласий по поводу того, что традиционные СУБД обнаруживают серьезные недостатки и ограничения при применении их к историческим данным. Однако существуют различные предложения по исправлению ситуации. Одни берут традиционные СУБД в качестве базиса и добавляют к ним некоторые инструментальные средства, расширяющие функциональные возможности этих систем и приспособляющие их к нуждам исторического исследования (см., например, SOCRATES, разрабатываемый Л.Брере на основе dBASE IV). Другие разрабатывают совершенно новое программное обеспечение специально для исторических приложений (KLEIO М.Таллера, WinCENS Я.Олдерволла). Хотя оба подхода имеют свои "за" и "против", второй кажется более многообещающим в применении к более сложным источникам или сложным задачам анализа.

Это станет ясно, если мы более подробно рассмотрим систему KLEIO. Стоит заметить, что эта система способна справиться с любыми типами "нерегулярности" данных, описанными выше. Но все же более важным

является то, каким именно способом система поддерживает интерпретацию исторических данных. Основной принцип построения системы KLEIO содержит двухуровневый подход: в то время как БД содержит материал исторического источника в близкой к оригиналу форме, система позволяет управлять "смыслом" элементов данных, а не текстовыми или числовыми знаками. Все выполняемые над данными операции, позволяющие нам извлекать, классифицировать или категоризовать данные и устанавливать связи между их элементами (имеются в виду индивидуальные связи), не ссылаются непосредственно на знаки или символы, содержащиеся в полях БД, а работают на своего рода смысловом уровне, выведенном из элементов данных.



В этой модели промежуточный программный слой, совмещающий различные инструменты, алгоритмы, лексику и т.д., представляет то, что может быть названо внешним историческим знанием /5/. Внешние знания используются для интерпретации данных: они содержат дополнительную информацию о данных и отражают предположения и гипотезы историков об их источниках. С помощью такого промежуточного слоя смысл, выводимый из элемента данных, используется для последующей обработки. Обе части – и данные, и внешние исторические знания – хранятся как самостоятельные компоненты БД; они обрабатываются совершенно независимо друг от друга и связываются вместе лишь при необходимости.

Стремление смоделировать процедуру конструирования смысла более адекватным образом привело к расширению промежуточного слоя реализацией так называемого контекстно-зависимого оперирования данными (КЗОД). Когда значение выводится из элемента данных, то его контекст принимается во внимание. Такой подход отражает довольно тривиальное положение: для интерпретации элемента данных не только сам элемент, но и его контекст может иметь критическое значение. Смысл текстового знака формируется динамически посредством интерпретации знака в его контексте. В терминах БД: значение атрибута должно быть "вычислено" с учетом контекстной информации, содержащейся в других атрибутах.

Преимущества КЗОД двойки: смысл может быть выведен из данных гораздо более адекватно, и в то же время обеспечивается большая гибкость на стадии ввода и возможность сохранить верность источнику. Как следствие, два элемента данных, идентичных по своим внешним характеристикам, могут быть по-разному обработаны системой, в зависимости от контекста, в котором они встретились. В экстремальных случаях КЗОД предполагает также обращение с элементом данных, содержащим строку символов "белый", как если бы он содержал термин "черный", если это обусловлено контекстом.

Подчеркнем, что этот подход довольно сильно контрастирует с базовым принципом обычных СУБД, а именно: содержание атрибутов (полей) в БД может быть доступно и адекватно обработано без учета значений других полей, и очевидно, что он имеет исключительное значение именно при создании БД в гуманитарных науках.

Хотя реализация контекстно-зависимых правил вывода включает некоторые методы ИИ и определенным образом отражает распространенную тенденцию в области БД /6/, в то же время мы должны очень четко представлять себе, что такие программные решения не имеют целью или все еще далеки от понимания языка машиной, т.е. от автоматической реконструкции смысла текста. В настоящий момент нас интересуют только некоторые фундаментальные положения и принципы, позволяющие моделировать на компьютере процедуру конструирования смысла, как это требуется для наших целей /7/.

Так, в соответствии с базовыми принципами созданная с помощью KLEIO БД содержит все данные как символьные строки, которые раскрывают свой смысл только в комбинации с историческими знаниями, так сказать генерируя контекст для понимания данных. Благодаря наличию промежуточного слоя не элемент данных сам по себе, а его "смысл" передается вызываемым функциям и используется для последующей обработки, обеспечивая, например, компьютерную поддержку приведения текстовой информации к четко определенным понятиям, как того требует количественный анализ. Однако этот базовый механизм имеет важные следствия для общего построения системы. Если мы всерьез выбираем такой подход и нацеливаемся на более адекватную поддержку исследовательского проекта, то внешние исторические знания, содержащие гипотезы и предположения о данных, должны быть организованы в форме, допускающей постоянное изменение: они могут часто пополняться, а во многих случаях и модифицироваться или полностью пересматриваться, если анализ данных приводит к необходимости новых представлений об отдельных их аспектах. Другими словами, связь между данными и знаниями не статична, а высоко динамична, данные и знания должны быть модифицируемыми на любом этапе работы. Это является причиной, по которой знания и данные должны быть не только организованы независимо друг от друга, но и быть связанными лишь в тот момент, когда данные

непосредственно требуются, т.к. смысл элемента данных должен отражать самое последнее состояние того, что называется "внешними историческими знаниями".

Важным следствием такого построения является то, что оценивание смысла элемента данных может происходить лишь по ходу работы, динамически (at run-time). Однако это означает, что всегда при выполнении операций над данными необходимо проделать много дополнительной работы: в любое время, когда осуществляется доступ к элементу данных, его смысл должен быть "вычислен" посредством подключения экспертного исторического знания и, если мы говорим о контекстно-зависимой обработке данных, контекста рассматриваемого элемента. Вся эта работа должна быть проделана в дополнение к стандартным операциям СУБД, замедляя их и в целом уменьшая производительность системы. Поэтому надо стараться минимизировать количество вычислений по ходу работы и убедиться, что поле переоценивается только тогда, когда это действительно необходимо /8/.

Мы приходим, таким образом, к достаточно передовой концепции в области технологии разработки БД - к концепции активной БД. СУБД должна гарантировать, что смысл элемента данных, переданный вызываемым функциям, отражает новейшее состояние внешнего знания, используемого для интерпретации. Хотя, с одной стороны, СУБД преследует цель уменьшить количество вычислений значений полей БД по ходу работы, с другой стороны, используемый для этого механизм должен гарантировать, что при обращении к элементу данных не используются устаревшие значения. Выведенные значения и заданные ограничения должны автоматически обрабатываться системой. Всегда, когда происходит обновление данных или внешних знаний, влияние этой модификации может распространяться - напрямую или косвенно - на большое число элементов данных; в любом случае все такие данные должны быть переоценены, для того чтобы быстро привести интерпретацию документов в соответствие наличной информации.

Более того, в описываемой СУБД проблема целостности данных также требует специального решения; здесь сравнение с другими системами едва ли возможно. Поскольку такие БД содержат данные как строки и присваивают смысл элементам данных в соответствии с внешними историческими знаниями, которые, по определению, постоянно изменяются, то трудно гарантировать целостность данных на этапе ввода. Во время заполнения БД смысл данных не может быть "вычислен", т.к. контекстная информация или внешние знания, требуемые для оценивания элемента данных, могут еще не быть доступными. Таким образом, очевидно, что выведенные из семантики данных ограничения не могут быть использованы для контроля целостности.

Во избежание ситуации, при которой данные вводятся в БД без какого-либо контроля над их целостностью - это означало бы

невозможность автоматического отсеивания дефектных данных и необходимость последующего поиска и исправления всех ошибок ввода – данные должны проверяться на соответствие синтаксическим ограничениям целостности, чтобы гарантировать по крайней мере один определенный аспект целостности. Ошибки же, нарушающие семантические ограничения целостности, выводимые из семантики данных, могут быть обнаружены только после того, как БД заполнена данными и укомплектована внешними знаниями, – в ходе работы над данными либо с помощью специальных программ.

Так как семантическая целостность может быть проверена лишь после ввода всех данных, то элементы данных, нарушающие ее, приходится рассматривать как часть БД и предоставлять на усмотрение пользователя решать, исправлять ли данные, изменять ли ограничения целостности или оставить все как есть. Значениям элементов данных, нарушающих семантические ограничения целостности, присваивается смысл (значение) "недействительный". Хотя очевидно, что такая БД допускает недействительные значения, их влияние должно находиться под контролем пользователя /9/.

Таким образом, для полной реализации описанной двухуровневой модели и обеспечения максимальной гибкости при обработке исторической информации в соответствии с целями ИООД требуются достаточно нетрадиционные решения в области технологии БД. Хотя и не все упомянутые аспекты уже реализованы в текущей версии КЛЕЮ, новые версии системы будут значительно развиты в этом направлении.

Базовая двухуровневая модель может быть расширена до трехуровневой включением еще одного типа данных, которым СУБД также должна оперировать – изображениями или битовыми картами /10/. Используя новые технические возможности, мы можем анализировать изображения, давно уже привлекающие внимание исследователей как самостоятельный тип источников, и на этой основе включать их в парадигму, называемую "источнико-ориентированной обработкой данных".

Ситуация становится еще более интересной, если обрабатываемое изображение есть не что иное, как изображение (образ) исторического текста, манускрипта. Здесь открываются новые горизонты применения компьютера для обработки исторических текстов. Это станет очевидным, если внимательнее взглянуть на фундаментальную проблему применения БД в историческом исследовании. Хотя ИООД направлена на возможно более полное и близкое к оригиналу представление источника, использование технологии создания БД в определенном смысле противоречило этой цели. Системы БД (за исключением полнотекстовых систем) требуют информации, вводимой в соответствии с определенной структурой или порядком, которые в большинстве случаев не присутствуют явно в источнике. Если мы имеем дело с текстовыми данными, то извлекаем информацию из источника, руководствуясь определенной логикой, которая определяется

требованиями СУБД. Но даже если в этом аспекте система позволяет достаточную гибкость, то часто нам приходится разрушать по меньшей мере оригинальные синтаксические формы текста. Другими словами, мы пытаемся сохранить смысл источника, но теряем его синтаксис. То есть почти неизбежно при вводе в компьютер информация теряется.

Для преодоления этого пробела были предложены некоторые решения, в том числе так называемая форма "подготовленного ввода"/11/, в соответствии с которой текст вводится в компьютер полностью, в то время как значимые фрагменты текста маркируются специальными символами, которые позволяют распознавать различные части текста и рассматривать выделенную информацию как материал для БД. Но включение образов текста в обработку исторических источников выглядит даже более перспективным. Если описывать это простыми словами, то имеется в виду следующее: хорошая репродукция источника конвертируется в оцифрованный образ текста; компьютер еще не знает текста источника, но хранит его как изображение (подобно фотографии текста). Когда текст представлен в такой форме, его можно легко превратить в машиночитаемый текст или ввести в БД. Затем с источником можно работать, пользуясь логикой СУБД, которая - как упоминалось прежде - предполагает некоторую структуру или порядок. И все, что мы извлекаем из источника, должно быть введено в поля БД и связано с "оригиналом" источника - точнее - с соответствующим фрагментом его образа. Это подводит нас к более широкому утверждению: оцифрованный образ источника или его фрагмента может быть связан с полями БД, содержащими описание источника в форме, близкой к оригиналу, и обрабатываемыми - как обычно - на основе внешнего исторического знания.

Наверное, стоит упомянуть, что включение изображения текстового источника в среду БД не только позволяет лучшим образом воспроизвести исторический документ, но и дает дополнительные возможности. Так, с использованием методов улучшения изображения можно сделать манускрипт более читаемым путем усиления контрастности, повышения чистоты и т.п. Другими словами, компьютер может помочь в работе над текстами, давая возможность использовать методы, которые обычно не могут быть приложены к исходному документу.

В систему KLEIO уже интегрированы средства управления изображениями как объектами БД, которые представляют собой вполне цельную подсистему. До настоящего времени обработка изображений в основном еще сильно зависит от мощных технических средств, то же самое относится и к обработке изображений модулями KLEIO. Однако благодаря постоянному понижению цен и быстрому развитию техники средства обработки изображений скоро станут доступны все большему числу историков. Таким образом обработка изображений станет более распространенной частью компьютерных приложений для исторических исследований. Несомненно, что управление изображениями как объектами

БД будет особенно полезно для ИООД, предоставляющей более адекватный способ представления исторического источника в машиночитаемом виде.

ПРИМЕЧАНИЯ.

1. См.: R.M. Bell, E.H.G. van Cauwenberghe: The Medieval and Early Modern Data Bank (MEMDB)// Tijdschrift voor Geschiedenis, N.103 (1990), p. 260-278.

2. M.Thaller: "Что такое "источникоориентированная обработка данных?" // История и компьютер. Под ред. Л.Бородкин, В.Леверманн. (Halbgraue Reihe zur Historischen Fachinformatik, Vol. A15), St. Katharinen 1993), p. 7.

3. Международная Ассоциация "История и компьютер" (АНС) недавно начала дискуссию по проблемам выработки критериев оценивания степени соответствия различных СУБД задачам исторического исследования.

4. Обобщающее название группы алгоритмов нечеткого (по созвучию) поиска текстовых данных. - Прим. перев.

5. Мы используем термин "знание" более многозначно, чем это делается обычно в сфере искусственного интеллекта (ИИ), где это понятие всегда связано с какой-либо процедурой логического вывода; многие элементы промежуточного слоя не могут рассматриваться как средства для представления "знания" в этом смысле.

Имеется в виду знание, внешнее по отношению к источнику. В оригинале - historical background knowledge. - Прим. перев.

6. Основной причиной интереса к взаимодействию между ИИ и БД является "осознание того, что с одной стороны, существенное увеличение продуктивности и функциональных возможностей информационных систем требует рассмотрения информации как з н а н и я и использование методов ИИ, таких как логический вывод, решающие правила, вопрос-ответные системы, предлагаемых экспертными системами. С другой стороны, практические приложения этой технологии, особенно те, что используют большие количества информации, требуют методов реализации, значительный прогресс в развитии которых достигнут в области БД (см. введение к M.L.Brodie and J.Mylopoulos in: On Knowledge Base Management Systems. Integrating Artificial Intelligence and Database Technologies. Ed. by M.L.Brodie, J.Mylopoulos. New York etc 1986, p. XI).

7. В связи с этим мы должны сделать два очень важных замечания: 1) Хотя сегодняшние семантические теории, особенно компьютерное моделирование понимания смысла, представление знаний и анализ текстов, обычно имеют дело с проблемой неоднозначности информации, эта тема встает со всей полнотой и станет объектом исследования в будущем. В течении некоторого времени полагалось, что знания, необходимые для развития контекстного подхода, могут быть четко формализованы и что

смысл также может принять некую форму и быть выражен тем же способом.
2) КЗОД реализована лишь как вариант KLEIO и не содержится в современной распространяемой версии программы.

8. В соответствии с базовой концепцией смысл элемента данных существует в БД лишь виртуально и должен выводиться всегда, когда осуществляется доступ к нему. Смысл элемента данных остается прежним, лишь до тех пор, пока сам элемент или внешнее знание, использованное для интерпретации, не дополняется. Другими словами, в течение того времени, когда изменения не вносятся, количество вычислений может быть уменьшено путем сохранения смысла элемента данных для того, чтобы иметь его "под рукой" для следующего раза, когда осуществляется доступ к этому элементу.

9. Для более детального обсуждения вопросов эффективности и целостности данных в такой системе см. W. Levermann: Context-Sensitive Reasoning in Historical Computing. Статья представлялась на VI-й Конференции Международной Ассоциации "История и компьютер", проходившей 28-30 августа 1991г. в Оденсе, Дания (в печати).

10. В русскоязычной литературе термин "битовая карта" используется как правило в узком смысле, для обозначения одного из форматов графических файлов. В англоязычной - аналогичный термин (bit-map) часто применяется шире - к любым введенным в компьютер изображениям. - Прим. перев.

11. В оригинале - pre-edited input. Для обозначения этой технологии подготовки машиночитаемых текстов чаще используется термин markup, markurping. - Прим. перев.

МЕТОДИКА ОЦЕНКИ ИНФОРМАЦИОННОГО ПОТЕНЦИАЛА
КОМПЛЕКСОВ МАССОВЫХ ИСТОЧНИКОВ

В.И.Тихонов, В.С.Тяжелникова, И.Ф.Юшин

Последние годы ознаменованы завершением работы по рассекречиванию крупных комплексов архивных документов, имеющих непреходящее значение для научного изучения советской истории. Немаловажным является также и то обстоятельство, что общее ухудшение экономической ситуации в стране повлияло на возможности финансирования научными учреждениями проектов, связанных с большими объемами ввода первичных данных. Вместе с тем открытые для исследователей комплексы массовых материалов настолько значительны по содержащейся в них информации и многомерны, что существующий уровень историографического осмысления сопряженных с ними проблем явно недостаточен. Характерным примером в этом отношении может служить комплекс личных дел жителей Москвы, ходатайствовавших о восстановлении в избирательных правах (далее – лишенцы), хранящийся в настоящее время в Центральном муниципальном архиве города /1/.

В этой ситуации задачами архивистов являются систематическая работа по привлечению внимания научной общественности к наиболее интересным новым комплексам массовых источников, демонстрация их информационных возможностей, определение круга проблем, который может быть изучен на их основе, а также, что хотелось бы особенно подчеркнуть, разработка методик введения таких источников в научный оборот на уровне системного анализа. В качестве практического результата подобных методик могут явиться рекомендации исследователям, изучающим конкретно-исторические проблемы, по оптимальной работе с большими по объему комплексами массовых материалов.

В настоящей работе предлагается один из вариантов введения комплекса массовых источников в научный оборот, основанный на демонстрации его информационных возможностей. Под этим мы будем понимать определение основных направлений конкретно-исторических исследований и оптимального пути их решения. Круг возможных конкретно-исторических тем определяется в результате компьютерного моделирования, а формирование информационных основ исследования по каждой теме предлагается исходя из количественных характеристик этой модели.

I. Информационный потенциал массового источника

Подход к источнику, существующий до настоящего времени, предполагает классический путь от постановки проблемы на основе

историографического анализа ее состояния до подбора источников, содержащих достаточные в качественном и количественном отношении сведения. Предлагаемая нами методика, ориентированная прежде всего на источник, базируется на выявлении его информационного потенциала. Таким образом выявляются те проблемы, для изучения которых информационный потенциал данного источника наиболее высок.

Под информационным потенциалом комплекса массовых источников мы понимаем наличие в нем достаточной информации для разработки конкретной научной проблемы. Вне проблемы, то есть при постановке всего спектра возможных содержательных задач, информационный потенциал комплекса тождествен всей его информации.

Информационный потенциал массива имеет качественные и количественные характеристики, выявление и измерение которых позволяет определить достаточность информации для проведения исследования.

Наша методика ориентирована на измерение информационного потенциала массовых источников при соответствующей их обработке в специальном программном обеспечении. Она базируется на построении компьютерной модели, и применение ее наиболее оптимально в отношении нарративных источников.

Определение информационного потенциала источника невозможно без систематического анализа как внешних факторов, влияющих на появление того или иного комплекса массовых материалов, так и главным образом его внутренней природы. Изучение логики расположения информации, составляющей "рисунок" источника, учет всей полноты контекста позволяют, на наш взгляд, реконструировать документ в форме компьютерной модели и руководствоваться результатами анализа небольшого по объему "пробного" массива для демонстрации информационного богатства источника, его возможностей как носителя знаний об эпохе.

Надо сказать, что массовые источники системно могут быть проанализированы только на основе современных технических средств, так как все остальные способы основаны на их выборочном, фрагментарном изучении и приводят к тому, что они изучаются как система отдельных сведений, или сведения агрегируются до такого состояния, когда специфика документа нивелируется. Внутренняя организация информации, взаимосвязи как иерархические, так и неиерархические остаются за гранью интерпретации, которую получает источник. На этом можно было бы и остановиться, если бы дальнейшее исследование, включающее зачастую построение сложных математико-статистических моделей, а также аналитические выводы автора не строилось именно на этой, в той или иной мере априори искаженной интерпретации. Такой подход до последнего времени был общепринят, поскольку не существовало необходимого программного обеспечения для того, чтобы корректно перенести все

нюансы документа в машиночитаемый вид и, создав модель, впоследствии сосредоточиться на ее анализе.

Появившиеся в последние 5-10 лет системы, ориентированные на специфику исторического источника и обработку исторических данных, отвечают, на наш взгляд, наиболее высоким требованиям, предъявляемым исследователем-источниковедом к гибкости программного обеспечения, и позволяют вводить сведения в компьютер в их "естественном" виде, при желании вовсе исключить формализацию. Точнее, речь идет о таком подходе, когда источник с его сложной и многомерной структурой описывается средствами языка компьютерной системы, а не наоборот - когда система диктует структуру модели. При создании такой модели определяющую роль играют два фактора: специфика источника и исследовательская задача. Это позволяет, сформулировав исследовательскую задачу как источниковедческий анализ конкретного комплекса массовых источников, полностью ориентироваться на специфику документа.

Таким образом, модель включает в себя описание структуры источника и его свойств в идеальном представлении исследователя с помощью программных средств компьютерной системы. Совокупность модели и соответствующих ей первичных данных источника составляет базу данных.

Выделение основных информационных блоков предполагает также их расположение в определенной иерархии, учитывая все многомерные связи, существующие между ними. Вместе с тем эти информационные блоки, которые в используемой системе определены как "информационные группы", не являются простым воспроизведением личного дела, а определены именно структурой информации, содержащейся в нем. Их определение является результатом аналитической работы исследователя по построению наиболее адекватной модели информационной структуры источника. Построение модели - это наиболее важный этап исследования, поскольку от корректности этого построения зависит результат решения поставленных проблем.

Создание компьютерной модели не является в конечном итоге самоцелью. В процессе работы происходит ее качественное насыщение, складывается идеальный вариант источника, вбирающий в себя все потенциально возможные элементы информации, присущие данному массиву. Обнаруженные в процессе создания компьютерной модели элементы информации интерпретируются и в смысле тех содержательных задач, которые могут быть изучены. Этот этап - важнейший шаг в качественном определении информационного потенциала источника. Наполнение этой модели при работе с выборкой позволяет количественно оценить информационный потенциал массива.

2. От источника к его компьютерной модели

Система КЛИО, на которую мы будем ориентироваться при нашем анализе, основана на нереляционной идеологии построения баз данных и

предполагает размещение информации в двух файлах - файле, содержащем первичные данные, и файле, содержащем модель данных, которая средствами командного языка КЛИО описывает структуру источника. Первоначально загружается модель, а после этого зачитываются первичные данные с помощью специальных управляющих символов. Такая организация позволяет редактировать и уточнять модель параллельно вводу данных, проверяя на практике и уточняя те особенности документа, которые не были однозначны и очевидны на первый взгляд.

Идентичной единицей базы данных КЛИО является документ, который включает специфическим образом организованные наиболее крупные блоки сведений - информационные группы с их наполнением - элементарными информациями. Эти положения наиболее важны в свете поставленной нами задачи. Заметим, что ориентация на исторические данные при создании системы привела к тому, что эквивалент поля - элементарная информация - имеет безграничную длину, позволяет включать в нее большое количество логически равноправных записей, а также содержать два дополнительных аспекта: примечание и фрагмент оригинального текста.

Остановимся более подробно на способах создания модели в КЛИО. В нашем случае в качестве отдельного документа выступает личное дело лишенца, а модель позволяет корректно расположить в рамках документа все имеющиеся наиболее крупные блоки сведений, перенести их не просто механически, а учесть всю совокупность многомерных связей, которые присутствуют в реальном источнике. Массовый характер личных дел предполагает наличие единых закономерностей организации информационных групп для всего массива в целом, а система позволяет на основе единой модели корректно расположить их в порядке, характерном для данного конкретного дела.

Идентичной единицей - документом - нашей базы данных является личное дело. Поэтому в центре модели и во главе иерархии будет располагаться информационная группа "дело" с соответствующими элементарными информациями - номер дела, опись, фонд, район. Таким образом, здесь закладывается реальное положение вещей: "нет персоналии без дела", нами не учитывается лишенец, на которого не было заведено дело.

С другой стороны, это выливается в то, что от группы "дело" зависит единственная группа - "персоналия", которая включает наиболее общие сведения о лице, подававшем заявление или жалобу в избирательную комиссию о восстановлении в правах. К этим сведениям мы относим такие, как фамилия, имя, отчество, адрес, дата рождения и др. Система предлагает при вводе данных следующую возможность: в файл заносятся только те сведения, которые реально присутствуют в деле. Отсутствие информации при вводе не обозначается. Это положение справедливо и для информационных групп. Анализ документов показывает, что в процессе обращения в комиссию лишенец доводил до ее сведения те или иные

подробности своей биографии. Эти сведения логично расположить в так называемых "биографических" информационных группах: "материальное положение", "образование", "занятие", "общественная работа", "прибытие в Москву". Для информационных групп справедливо то же правило, что и для элементарных информаций: если, например, сведения для группы "материальное положение" в деле отсутствуют, то при вводе данных эта группа просто опускается.

Принципиальным моментом, с точки зрения содержания документа и его отражения в модели данных, является наличие в деле нескольких субъектов, которые в той или иной мере связаны с обстоятельствами биографии лишенца или выступают в роли носителей определенной информации. Одним из таких, наиболее часто встречающихся субъектов является родственник лишенца. Родственные отношения очень часто были причиной лишения прав: по существующим нормативным актам избирательных прав лишались лица, жившие на средства "нетрудовых и эксплуататорских элементов", что соответствующим образом отразилось в источнике. Поэтому справедливо, на наш взгляд, включение в модель информационной группы "родственник", которая зависит от группы "персоналия".

Система позволяет в случае необходимости включать в модель несколько логических уровней, которые помогают обозначить взаимосвязи. На схеме I видно, что те же информационные группы, которые описывают персоналию и зависят от нее, подчиняются и информационной группе "родственник", это отражает реальную ситуацию: в деле, когда лишение избирательных прав излагается в связи с родственными отношениями, биография родственника отражена столь же подробно, как и биография самого лишенца. Система четко отличает по последовательности расположения групп в первичных данных, применительно к какому субъекту дела их следует интерпретировать. Так, например, если мы вводим последовательность "персоналия-занятие", то КЛЮ будет ее интерпретировать как сведения о занятии лишенца. Если же имеется последовательность "родственник-занятие", то при обработке сведений это будет трактоваться как занятие родственника.

Вторым типичным действующим лицом очень часто являлся свидетель. Он вызывался в комиссию или представлял те или иные сведения, о которых не было других подтверждений. Что это был за человек и каким образом его биография была связана с жизненным путем лишенца, немаловажно, поскольку очень часто от этого зависит характер информации, полученной с его слов. Иногда свидетелей привлекал сам лишенец, в других случаях ими были авторы доносов.

С точки зрения создания компьютерной модели личного дела весьма существенную роль играет фиксация источников тех или иных сведений. От этого очень часто зависит корректность интерпретации того или иного факта. В структуру модели личного дела мы включили две информационные группы, сведения которых имеют более источниковедческое, чем чисто

историческое значение. Это информационные группы "личные документы" и "сопроводительные документы". С целью облегчения последующей обработки данных мы намеренно классифицировали документы дела по их авторству. Для нас в этой связи было важно, был ли лишенец автором того или иного документа или нет (если был, то сведения о нем входят в группу "личный документ", если нет, тогда они включаются в группу "сопроводительный документ"). Такая очень гибкая классификация документов носит в известном смысле вспомогательный характер и позволяет впоследствии более подробно классифицировать все справки, содержащиеся в личных делах. Как видно из схемы, дело могло содержать личные и сопроводительные документы как в отношении самой персоналии, так и в отношении родственников.

Наиболее существенным вопросом в контексте лишения-восстановления в избирательных правах был вопрос о собственности. Именно в этой группе, как показывает анализ личных дел, наиболее четко прослеживается зависимость сведений от источника. Именно по вопросу о собственности сведения, которые давал сам лишенец (в документах личного происхождения), имеют наибольшие отличия от сведений документов, определенных нами как сопроводительные. Это отражается в модели следующим образом: последовательность "персон - ЛД - собственность" система интерпретирует как сведения о собственности по личным документам, последовательность "персоналия - СД - собственность" как сведения о собственности по сопроводительным документам. Обе последовательности имеют отношение к собственности самого лишенца. Реже встречаются сведения о собственности родственников и свидетелей, расхождения в этом случае достаточно будет установить по номеру листа, поэтому группа "собственность" в модели будет непосредственно подчиняться группам "родственник" и "свидетель".

Анализ личных дел показывает, что существенное место в структуре источника занимает бюрократическая процедура лишения-восстановления в избирательных правах, которая по своему характеру может быть квалифицирована как судебно-следственная. В процессе лишения избирательных прав и процедуре восстановления можно выделить три самостоятельных логических эпизода - собственно факт лишения, последовавшие за ним заведение дела и его рассмотрение в комиссии, окончательное решение по нему. Поэтому характер процедуры и ее отражение в источнике позволяют выделить три информационные группы - "лишение", "ход дела", "окончательное решение". Первая группа здесь имеет центральное значение, поскольку сама бюрократическая процедура становилась возможной только после факта лишения. Следовательно, две последние группы в нашей структуре зависят от нее. Иными словами, мы отражаем в структуре базы данных следующий факт: "без лишения прав невозможны ход дела и принятие по нему окончательного решения". Группа "ход дела" имеет наибольшую степень логического обобщения. В личных

делах довольно редко встречаются прямые указания на те бюрократические этапы, которое должно было пройти дело в ходе своего рассмотрения, прежде чем оно отправилось в архив. Поэтому исследователь должен самостоятельно разобраться во всех процедурных вопросах, в логике перемещения дела лишенца по инстанциям и зафиксировать те этапы рассмотрения дела, которые нашли отражение в документах. "Окончательное решение" - фиксация того факта, что рассмотрение дела закончилось принятием определенного решения, которое, судя по материалам архивного дела, может считаться окончательным в данную избирательную кампанию. То есть нет оснований судить о том, что лишенец обращался далее с новой жалобой в вышестоящую инстанцию. Таким образом, мы реконструируем отразившуюся в источнике сложную и запутанную процедуру, сохраняя указания на расположение информации о ней в документе.

Большинство групп в нашей модели не являются механическим воспроизведением личного дела, а позволяют реконструировать основные блоки информации в нем с учетом взаимосвязей между ними. Определение информационных групп может диктоваться исследовательской задачей, которая в данном случае сводилась к тому, чтобы максимально адекватно интерпретировать источник - личное дело лишенца - в форме компьютерной модели.

Построив компьютерную модель массива и проанализировав его информационную структуру, мы можем предложить исследователям следующие направления исторического поиска: социальный портрет лишенца, его близких и окружения; процедура восстановления в избирательных правах и ее социальные аспекты; эволюция и формирование социальных групп столицы; моделирование социального поведения граждан в условиях формирования административной системы управления советского государства.

Определение основных направлений исследовательского поиска лишь первый этап в демонстрации информационного потенциала комплекса массовых источников и введения его в научный оборот. Следующим шагом должно стать определение оптимальных условий работы с источником через анализ количественных характеристик выявленных в нем элементов информации.

3. От модели - к оценке информационного потенциала источника

При работе с массовыми источниками над любой проблемой совершенно не обязательно использовать весь комплекс целиком, достаточно какой-то репрезентативно отобранной его части. Работа с массивом личных дел лишенцев определяется следующими его особенностями: различной частотой встречаемости сведений определенного характера, их фрагментарностью. Это приводит к тому, что для работы над разными конкретно-

историческими проблемами требуется различный объем выборки. Определение объема репрезентативной выборки для работы с каждой информационной группой (или их совокупностью) позволяет оценить количественные характеристики информационного потенциала комплекса применительно к конкретной содержательной задаче.

Определив информационную базу для решения поставленной задачи, исследователь далее работает только с этими информационными группами. Генеральной совокупностью для него является общий объем этих групп, содержащийся в базе данных. Проблема в том, что в таких слабоструктурированных источниках, как личные дела, величина этого объема исследователю неизвестна. Определить этот объем исследователь может двумя способами: либо просмотрев все дела массива, либо на основе экспериментально установленных количественных характеристик оценить размер генеральной совокупности по репрезентативному количеству дел. Второй способ, хотя и менее точен, зато гораздо менее трудоемок. Зная среднее количество информационных групп в документе, исследователь может рассчитать как общее количество соответствующих групп в генеральной совокупности, так и объем личных дел, содержащих репрезентативное количество групп. Предлагаемый подход, основанный на определении количественных характеристик компьютерной модели источника, помогает исследователю выяснить, сколько личных дел необходимо, чтобы с достаточной достоверностью и точностью судить о размере генеральной совокупности по каждой информационной группе.

Из всех хранящихся в ЦМAM фондов с личными делами лишенцев мы случайным образом отобрали один, содержащий 3541 личное дело /2/. В процессе создания компьютерной модели нами была сформирована 3%-ная экспериментальная выборка (105 дел), которая позволила уточнить качественные параметры модели. Следующим шагом явилась оценка надежности этой выборки для характеристики генеральной совокупности. Для этого мы на основе экспериментальной базы данных сформировали таблицу, где в качестве показателей использовались информационные группы нашей модели, а в качестве объектов - личные дела. За значение показателя мы принимали количество информационных групп в каждом из 105 дел. Расчеты предельной ошибки экспериментальной выборки продемонстрировали, что по всем показателям (при доверительной вероятности 0,9) эта ошибка составляет более 10% от среднего значения количества информационных групп. Это довольно значительная величина. При определении объема репрезентативной выборки мы ориентировались на то, что предельная ошибка не должна превышать 10% от среднего значения признака.

Расчет объема репрезентативной выборки для количественного значения показателя проводится по следующей формуле:

$$n_1 = \frac{N * t^2 * \sigma_0^2}{N * d_1^2 + t^2 * \sigma_0^2} \quad (1),$$

где n_1 - объем репрезентативной выборки; N - объем генеральной совокупности личных дел; σ_0^2 - дисперсия количественного признака в экспериментальной выборке; t - табулированное значение доверительной вероятности; d_1 - предельная ошибка выборки. Наши расчеты покажут, какое количество объектов (дел) нужно отобрать из генеральной совокупности, чтобы с определенной доверительной вероятностью утверждать, что в выборке среднее количество информационных групп по каждому объекту не будет отличаться от среднего количества информационных групп по каждому объекту в генеральной совокупности более чем на принятое значение d_1 .

Как видно из формулы 1, определение объема репрезентативной выборки зависит: 1) от объема генеральной совокупности - массива личных дел; 2) от распределения значений признака; 3) от уровня доверительной вероятности; 4) от заданной предельной ошибки выборки. Однако если объем генеральной совокупности известен заранее, а распределение значений признаков мы выясняем при помощи проведения экспериментальной выборки, то доверительная вероятность и предельная ошибка задаются самим исследователем. Чем выше вероятность или меньше ошибка, тем больше должен быть объем репрезентативной выборки. Следует иметь в виду, что слишком завышенные требования к выборке могут вообще поставить под сомнение необходимость ее проведения, так как полной достоверности и точности результатов исследования можно добиться только, если работать со всей генеральной совокупностью. Для наших расчетов мы принимали, что вероятность равна 0,9, а предельная ошибка выборки, как было сказано выше, - 10% от среднего значения количества информационных групп. Это означает, что в 9 выборках из 10 рассчитанного объема ошибка в определении среднего значения информационных групп не будет превышать более 10% среднего значения соответствующего показателя в генеральной совокупности.

Зная среднее значение показателя во всем массиве, исследователь сможет узнать объем генеральной совокупности по каждой информационной группе. Он также сможет выяснить, достаточна ли проведенная выборка для решения конкретной исторической задачи или нет. В этом случае исследователь имеет дело с качественными значениями показателей, если его интересует, например, социальный или профессиональный состав лиценцев. Здесь исследователь будет работать с качественными значениями признака, а расчетная формула примет вид:

$$n_i = \frac{N_i * t^2 * s^2}{N_i * d_2^2 + t^2 * s^2} \quad (2),$$

где n_i - объем репрезентативной выборки из объектов с признаком i (информационных групп); N_i - объем объектов с признаком i в генеральной совокупности (информационных групп); $s^2 = q_i * (1 - q_i)$ - дисперсия качественного значения признака i , где q_i - доля качественного значения признака; t - табулированное значение доверительной вероятности; d_2 - предельная ошибка выборки при определении доли качественного значения признака.

Достаточный объем выборки для работы с качественными значениями показателей мы попытаемся оценить по экспериментальной выборке.

Так как

$$N_i = N * x \quad (3),$$

$$n_i = n * x \quad (4),$$

где x - среднее количество информационных групп в генеральной совокупности, то

$$n_2 = \frac{N * t^2 * s^2}{N * d_2^2 * x + t^2 * s^2} \quad (5),$$

где n_2 - репрезентативный объем выборки при работе с информационными группами.

При этом необходимо учитывать, что доля того или иного качественного значения показателя нам неизвестна. Поэтому оценку объема выборки мы будем проводить по максимально возможной дисперсии этой доли: $s^2 = 0,25$, при $q_i = 0,5$. Кроме того, точное значение x нам тоже неизвестно, а так как величина n_2 зависит от величины x , то при оценке выборки лучше будет взять минимально возможное значение x . То есть

$$x = x_0 - d_1 \quad (6),$$

где x_0 - среднее количество информационных групп в экспериментальной выборке. Поэтому формула 5 принимает вид:

$$n_2 = \frac{N * t^2 * s^2}{N * d_2^2 * (x_0 - d_1) + t^2 * s^2} \quad (7),$$

При определении n_2 принимаем, что доверительная вероятность - 0,9, d_1 - 10% от среднего значения количества информационных групп в экспериментальной выборке, а $d_2 = 0,05$. Расчеты покажут, в каком объеме выборки с доверительной вероятностью 0,9 максимально возможное распределение качественного значения признака не будет отклоняться от такого же распределения в генеральной совокупности более чем на

значение d_2 . Иначе говоря, данный объем выборки вполне достаточен для работы с конкретной информационной группой.

Расчеты по формулам 1 и 7 показаны на рисунке 1. Как видно, объемы выборок n_1 и n_2 , при одной и той же достоверности и ошибке d_1 , сильно варьируются. Разница между величинами выборок зависит от многих факторов, но в первую очередь от величины d_1 . Чем точнее мы хотим определить среднее значение количества информационных групп в генеральной совокупности (а значит, чем больше n_1), тем меньше, в итоге, максимально возможное n_2 . Для исследователя количественное определение этой разницы может иметь далеко не праздный интерес.

Выяснив по формуле 1 объем репрезентативной выборки для определения среднего количества информационных групп в документе, следует определить, достаточен ли будет данный объем выборки для дальнейшей исследовательской работы, т.е. будет ли n_2 больше или меньше n_1 . Выяснить это можно только при определении отношения n_1 к n_2 :

$$\frac{n_1}{n_2} = k \tag{8}$$

где k - отношение объема выборки для определения среднего значения количества информационных групп к объему выборки по доле качественного значения признака.

Подставив в это выражение формулы 1 и 7, найдем, что

$$k = \frac{N \cdot d_2^2 \cdot x_0^2 \cdot (x_0 - d_1) + t^2 \cdot s^2 \cdot \sigma_0^2}{N \cdot d_1^2 \cdot s^2 + t^2 \cdot s^2 \cdot \sigma_0^2} \tag{9}$$

Таким образом, можно заранее выяснить значение k при любых заданных значениях t , d_1 и d_2 . Если $k > 1$, то объем выборки для определения среднего количества информационных групп будет достаточен для дальнейшей работы. Если $k < 1$, то выборку необходимо будет дополнить.

Конечно, идеальный вариант, когда $k=1$, т.е. когда $n_1=n_2$. Можно определить условия, при которых будет выполняться это равенство. При заданных значениях t и d_2 основным фактором, влияющим на величины n_1 и n_2 , является предельная ошибка d_1 в оценке среднего значения количественного показателя x в массиве дел. Определив оптимальную для нас предельную ошибку среднего значения d_r , мы и выясним оптимальную величину репрезентативной выборки для работы с конкретной информационной группой.

Если $k=1$, при $d_1=d_r$, то выполняется равенство:

$$\frac{N \cdot d_2^2 \cdot \sigma_0^2 \cdot (x_0 - d_r) + t^2 \cdot s^2 \cdot \sigma_0^2}{N \cdot d_r^2 \cdot s^2 + t^2 \cdot s^2 \cdot \sigma_0^2} = 1.$$

Отсюда:

$$N \cdot d_2^2 \cdot \sigma_0^2 \cdot x_0 - N \cdot d_2^2 \cdot \sigma_0^2 \cdot d_r - N \cdot d_r^2 \cdot s^2 = 0.$$

Решая данное уравнение, находим, что:

$$d_r = \frac{(N \cdot d_2^2 \cdot \sigma_0^2)^2 + 4 \cdot N^2 \cdot d_2^2 \cdot \sigma_0^2 \cdot s^2 \cdot x_0 - N \cdot d_2^2 \cdot \sigma_0^2}{2 \cdot N \cdot s^2} \quad (10).$$

Рассчитав значение d_r по каждому признаку и подставив его в формулу 1 или 7 (при $d_1 = d_r$), мы найдем оптимальный объем выборки для работы с любой информационной группой нашей базы данных. При расчете этого объема нами учитываются характерные особенности личных дел: включение в формулу 10 величины x_0 позволяет учитывать количество информационных групп, а включение σ_0^2 - их наличие или отсутствие в документе.

На рисунке 2 представлены результаты расчета выборок из комплекса личных дел (при одних и тех же значениях s^2 , t и d_2). Из рисунка видно, для работы с какими информационными группами репрезентативная выборка является минимальной. Так, для работы с информационными группами, отражающими процедуру лишения-восстановления, состав документов в деле, основные биографические сведения, материальное положение и занятие лишенцев, миграцию населения и родственные отношения, достаточно 10-15% выборки. Для работы с информацией о собственности лишенца, его образовании и общественной работе, о занятиях его родственников и для характеристики свидетелей достаточно 20-25% массива личных дел. В 50%-ную выборку укладываются информационные группы: "лишение", "сопроводительные документы", "материальное положение", "общественная работа" и "собственность" родственников лишенцев. Наконец, для того чтобы работать с остальной информацией источника ("ход дела", "окончательное решение", "личный документ", "образование" родственника и "собственность" свидетеля), необходимо будет брать более половины всей генеральной совокупности личных дел.

Таким образом, чем меньше объем необходимой репрезентативной выборки, тем выше, в нашем понимании, информационный потенциал комплекса массовых источников по данному показателю.

4. От определения информационного потенциала источника - к практической работе с комплексами массовых источников

Предложенная методика может рассматриваться в двух основных аспектах: в источниковедческом плане - с точки зрения фундаментального изучения специфики массовых источников и в практическом - исходя из конкретных вопросов, возникающих при приеме на хранение в архивы массовых документов. В обоих случаях методика может стать ключом к определению корректных условий работы с массивом, когда объективные условия ограничивают возможность его использования в полном объеме. В первом случае эти ограничения связаны с тем, что исследователь, как правило, работает только с частью комплекса массовых источников, во втором - что принять на хранение всю генеральную совокупность документов невозможно. Остановимся более подробно на возможном практическом приложении данной методики.

Ценность документа, необходимость его хранения можно определять с точки зрения роли его автора в историческом процессе или исторического значения самого документа. Можно руководствоваться практической значимостью источника для решения вопросов управления, собственности или его художественной ценностью. То есть оценивать документ и необходимость его хранения исходя из совершенно различных критериев.

Названные критерии не могут в полной мере и однозначно переноситься на проблему приема на хранение массовых источников, которая становится все более актуальной, и ее актуальность будет возрастать в ближайшем будущем. Очевидно, что существующая практика, в основе которой лежит опыт специалистов по отбору материала и формированию выборок, не может считаться удовлетворительной. Для полного хранения нужны слишком большие объемы хранилищ, а произвольный отбор ведет зачастую к одностороннему отражению исторического процесса, функционирование различных государственных и общественных институтов вытесняет жизнь реальных людей.

Чтобы выстроить стратегию приема источников на хранение, необходимы совершенно новые принципы определения ценности комплексов массовых источников, критерии формирования выборки. В основу такой системы, на наш взгляд, может быть положена методика, позволяющая качественно определить и измерить информационный потенциал массива. Именно информационный потенциал и является, на наш взгляд, наиболее важным и существенным элементом массового источника, критерием определения его ценности. Информационный потенциал с точки зрения адекватности отраженного в нем прошлого, с одной стороны, и возможности решения на его основе исторических проблем, с другой стороны, должен стать точкой отсчета при определении критериев приема на хранение таких документов, определения количественных параметров выборки.

Такой подход, на наш взгляд, может лежать в основе источниковедческой экспертизы массовой документации. Такая экспертиза может идти по двум направлениям: первый – выявление основных информационных групп документа, их характерных сочетаний применительно к различным видам источников. На основе определения информационного потенциала можно проанализировать обязательную номенклатуру документов, принимаемых на государственное хранение. Это позволит определить их возможности для изучения конкретных исторических проблем и устранил ненужное дублирование. Ведь уже сейчас можно сказать, что повседневная жизнь остается за рамками архивохранилищ, а управленческая документация сама по себе не может отразить реалии жизни. Поэтому правомерна постановка вопроса о проведении экспертизы списка массовых источников, в основу которой будет положен анализ информационных возможностей. И здесь предложенная методика может быть одним из возможных вариантов.

Вторым направлением источниковедческой экспертизы массовых источников может стать специальное исследование в отношении таких устойчивых, формализованных документов, как личный листок по учету кадров. Может быть разработан интегрированный показатель оценки информационного потенциала на основе серии экспериментальных исследований, учитывающих временные, региональные, профессиональные различия делопроизводства. Настоящая методика сохранит свою актуальность и в дальнейшем, поскольку переход к приему документов в машиночитаемом виде в принципе не снимает проблему отбора информации как при историческом исследовании, так и при решении вопросов организации архивного хранения.

ПРИМЕЧАНИЯ

1. См. ЦМАМ, ф.3031, ф.3094, ф.3095, ф.3096, ф.3099, ф.3103, ф.3108, ф.3109, ф.3118, ф.3119, ф.3120, ф.3121, ф.3122, ф.3123.

2. ЦМАМ, ф.3109, Комиссия по рассмотрению жалоб лиц, лишенных избирательных прав, при Коминтерновском районном Совете.

РОССИЙСКИЕ ИСТОРИКИ XVIII - начала XX вв.:
ПРОЕКТ И ИНФОРМАЦИОННАЯ СИСТЕМА

Д.А. Гутнов, В.А. Перевертень

Введение

В последние годы заметно возрос интерес историков к научно-биографическим исследованиям. Среди зарубежных исследователей эта область исторической науки даже получила специальное название *просопография*. По-видимому, нет необходимости доказывать научную и социальную ценность просопографического исследования и корпуса российских историков за период с начала XVIII в. по начало XX в. За указанный двухсотлетний отрезок времени число лиц, тем или иным образом причастных к российской исторической науке оценивается экспертами в несколько тысяч. Однако в сферу научно-биографических исследований попадают в основном представители, так сказать, элиты исторической мысли. Средний и низший слои этого айсберга из-за своей толщи до сих пор остаются практически за пределами внимания исследователей. В то же время очевидно, что получить достаточно адекватный динамический "групповой портрет" российских историков без рассмотрения всего их корпуса невозможно. Как и невозможно решать многие другие историографические задачи, требующие глобального пространственно-временного охвата множества научно-социальных характеристик историков. Наблюдающаяся ограниченность в этой области отечественной просопографии объясняется теми объемами информации, которыми необходимо оперировать как при построении "коллективной биографии" достаточно большой группы историков, так и при решении иных проблем комплексного, системного ее изучения. Справиться с подобными родами задачами, используя традиционную, так называемую "карандашно-бумажную" технологию за разумное время невозможно. Заметно расширить информационные, познавательные и коммуникационные возможности историографов, радикально облегчить и интенсифицировать их умственную деятельность, повысить ее комфортность можно лишь путем внедрения в процесс исторического исследования и обеспечивающие его сферы современных компьютерных технологий. Организовав информацию о российских историках в форме базы (банка) данных, мы обеспечим не только комфортные условия для получения необходимых частных или обобщенных сведений, но и дадим практическую возможность проводить исследования, о которых раньше можно было говорить лишь теоретически.

Появление в последнее время на рабочих столах отечественных историков персональных компьютеров сделало реальной постановку вопроса

о реализации компьютеризированного проекта "Российские историки XVIII - начала XX вв."

В данной статье описывается проблематика проекта научно-биографических исследований корпуса российских историков периода XVIII - начала XX вв., предлагается парадигма построения и типовая архитектура информационных систем для исторических исследований, на примере текущей (на сентябрь 1993 г.) версии информационной системы "Просис", предназначенной для поддержки исследований в рамках упомянутого проекта, излагаются принципы построения и реализации нового класса информационных систем - просопографических, даются рекомендации по осуществлению проекта с использованием системы "Просис".

Российские историки XVIII - начала XX вв.:
проблематика проекта.

Идея проекта "Российские историки XVIII - начала XX вв." возникла в конце 80-х гг. Кроме авторов этой статьи его инициатором является также профессор РГГУ В.А.Муравьев.

Проект преследует две главные цели, которые в настоящий момент можно обозначить следующим образом.

1. Систематизация и концентрация в единой компьютерной среде данных о научно-социальном облике всех, кто оказал влияние на становление и развитие российской исторической науки в период с начала XVIII в. по начало XX в., сведения о которых разбросаны по множеству разнообразных источников. Будучи сосредоточенными в одном месте, эти данные через соответствующие электронные средства распространения информации в то же время станут доступными широчайшему кругу пользователей.

2. Создание компьютеризированной информационно-операционно-коммуникационной основы для решения задач просопографического характера, требующих глобального пространственно-временного охвата множества характеристик тех, кого мы причисляем к российским историкам. К таким задачам можно отнести:

- различного рода развернутые обобщенные характеристики научно-социальных процессов, наблюдающихся в коллективном облике российских историков;
- типологизацию корпуса российских историков по различным показателям;
- построение двухсотлетней "родословной" представителей российской исторической науки;
- выявление динамики "коллективной биографии" российских историков;
- биогеографические исследования;

- поиск статистических закономерностей;
- гипотетическое восстановление недостающих сведений и др.

Проблемы, связанные с достижением целей проекта, условно можно разделить на две группы. Одну группу составляют историко-исследовательские проблемы, а другую - историко-технологические.

Центральной из историко-исследовательских проблем следует считать определение совокупности научно-социальных характеристик персоналий российской исторической науки. Часть из этих характеристик носит объективизированный характер и в принципе присуща любому историко-биографическому исследованию. Другая часть отражает специфические цели поставленных исследовательских задач и зависит также от таких факторов, как интересующий нас период, регион и контингент. Мы также отдаем себе отчет в том, что с расширением и уточнением круга задач первоначально выбранный набор характеристик будет корректироваться и дополняться.

Ниже в качестве примера мы предлагаем стартовый набор объективизированных характеристик, сгруппированный по смысловой близости в следующих разделах:

- базовые биографические данные, под которыми подразумевается информация о фамилии, имени, отчестве, годах рождения и смерти, времени начала и конца творческой деятельности;
- социальные характеристики, такие, как национальное происхождение, вероисповедание, языки творчества, социальное происхождение, достигнутое социальное положение, титул, высший достигнутый чин, полученные ордена и владение собственностью;
- образование, куда входят сведения, отражающие где и когда историк получил домашнее, среднее, высшее образование, каких профессоров слушал, а также получал ли второе высшее образование;
- служебная деятельность, который содержит информацию о виде деятельности, наименовании учреждения, где ученый работал, в какой должности и в какой промежуток времени;
- научная подготовка, в который включены данные о темах и годах написания квалификационных работ, начиная с дипломной работы и кончая докторской диссертацией, с указанием учебного или научного заведения, где работа была защищена, а также сведения о странах, учреждениях и времени прохождения зарубежных научных стажировок;
- научный рост, - дополняющий предыдущий раздел и содержащий данные о том, где и когда историк получал педагогические звания: приват-доцента, экстраординарного и ординарного профессора;
- общественно-политическая деятельность, который включает сведения о том, каким видом деятельности занимался ученый, в каких организациях и в какое время;

- философская ориентация, - заключающий в себе информацию о том, к каким философско-методологическим системам был привержен ученый в течение своей жизни;

- характер научной деятельности, - фиксирующий те области научной, архивной, публикаторской, педагогической, публицистической, меценатской и др. деятельности, в которых работал исследователь;

- сфера занятий историка, - включающий в себя сведения об основных областях исторической науки, в которых работал историк (всемирная история, российская история, археология, этнография, источниковедение, историография, архивоведение, востоковедение и т.д.), а также данные о том, какими регионами и периодами он интересовался;

- проблематика научной деятельности, - отражающий информацию о том, к какой области исторической проблематики тяготел исследователь: общеисторическим вопросам, истории экономики, права, общественной мысли, учреждений, внешней или внутренней политики, армии и флота, истории культуры или созданию исторических портретов;

- основные работы по истории, - хранящий библиографические данные о главных трудах ученого;

- научные связи - работы и мысли каких исследователей в России и за рубежом предпочитал историк и кто были его ученики;

- признание научных заслуг, - содержащий сведения о членстве исследователя в Петербургской Академии наук с указанием ученого звания и года его получения, в зарубежных академиях и научных обществах или об иных формах научных отличий.

Второй по значимости и первой по трудоемкости является проблема сбора сведений о представителях российской исторической науки прошлого. Ее сложность заключается в том, что не существует какого-либо унифицированного исторического источника, содержащего исчерпывающие данные обо всех историках. Информация о них находится в самых различных архивах, библиотеках, музеях и разбросана по самым разным публикациям. Различна и степень достоверности этих сведений, оценить которую зачастую может только специалист.

Нет необходимости доказывать, что залогом успешного решения очерченных историко-исследовательских задач является правильный выбор методов. Проще всего эта проблема решается в том случае, когда метод уже известен и хорошо себя зарекомендовал. Положение существенно усложняется для задач, требующих разработки новых методов. Следует подчеркнуть, что большинство просопографических задач относится как раз ко второй группе.

Главной историко-технологической проблемой является разработка соответствующей информационной системы, которая обеспечила бы как организованное хранение в компьютерной среде всей относящейся к проекту информации, так и компьютеризированное решение поставленных

историко-исследовательских задач. Среди факторов, затрудняющих решение этой проблемы, можно выделить следующие:

- отсутствие концептуально обоснованной типовой архитектуры информационных систем, предназначенных для исторических исследований;
- разнородность исторической информации;
- специфичность научно-познавательной деятельности историков и их менталитета;
- отсутствие теоретического базиса и практического опыта построения информационных систем для научно-биографических исследований.

Исторические информационные системы: парадигма разработки и архитектура

Среди компьютерных информационных систем те, которые специально созданы для поддержки информационно-познавательной деятельности историков, можно выделить в отдельный класс и назвать его "историко-исследовательские информационные системы" или "исторические информационные системы".

Специалисты по применению компьютеров в исторической науке отмечают, что создание информационных систем для исторических исследований является более сложной и специальной задачей, чем построение традиционных информационных систем (например для деловых приложений) /1/. Прежде всего, это связано с разнородностью историко-исследовательской информации, слабой, как правило, определенностью как постановки исследовательских задач, так и путей их решения, а также специфичностью научной деятельности историка.

Наблюдающиеся парадигмы использования компьютеров в истории в основном определяются ориентацией относительно исторических исследований и взгляда на применение компьютерных технологий /2/. В зависимости от отношения к информации источников, задачам и методам их решения называют следующие подходы:

- методо-ориентированный, когда совокупность данных, извлекаемых из источников, и форма их представления в компьютере определяются методом решения задач;
- источник-ориентированный, при котором делается попытка смоделировать на компьютере все множество содержащейся в источниках информации безотносительно к предполагаемым исследовательским задачам и методам их решения;
- проблемно-ориентированный, когда представляемая в компьютере источниковая информация отбирается в соответствии с общим видением поставленной проблемы и исходя из вытекающих из этого задач. Что касается применения компьютерных технологий в исторических исследованиях, то имеется три позиции /3/.

1. "Силовая" и наивная, игнорирующая специфику исторической информатики и опирающаяся лишь на стандартное программное обеспечение.

2. Осознающая фундаментальные отличия историко-исследовательской информации от "обычной" и ратующая за создание специализированного программного обеспечения вместо использования стандартного.

3. Промежуточная между двумя названными и выступающая за решение историко-исследовательских проблем на базе стандартного программного обеспечения, но специально доработанного.

К сожалению, в современной парадигматике компьютеризации исторических исследований не уделяется специального внимания одному из важнейших, по нашему мнению, вопросов, а именно отношению используемой или создаваемой компьютерной системы к историку как познающему субъекту. Опыт применения компьютеров в других областях (например в сфере управления) показывает, что успех в использовании некоторой компьютерной системы зависит не только от ее возможностей, но и от комфортности для пользователя. Что касается внедрения компьютеров в историческую науку, то в силу специфического, гуманитарного менталитета историков пренебрежение психологическими аспектами их информационно-познавательной деятельности является одной из существенных причин замедленности массовой компьютеризации этой отрасли познания.

В доступной нам литературе /4/ не удалось обнаружить работ, специально посвященных психологии познавательной деятельности историка. Поэтому выдвигаемый нами так называемый когнитивно-ориентированный подход к разработке исторических информационных систем, при котором максимально учитываются познавательно-психологические особенности пользователя-историка, основывается на имеющихся знаниях из когнитивной психологии, бытующих представлениях о технике исторического исследования, собственном опыте и интуитивном понимании проблемы и путей ее разрешения.

Парадигма, которой мы придерживаемся, предлагая типовую архитектуру исторических информационных систем и принципы создания таковых для просопографических исследований, заключается в проблемно-ориентированном подходе к компьютеризации, когнитивно-ориентированном подходе к построению компьютерных систем и разумном сочетании стандартного программного обеспечения со специально разработанным.

В основе описываемой ниже архитектуры исторических информационных систем кроме концепций построения общих для всех современных информационных систем лежит представление о том, что познавательные процессы протекают в определенной среде, которую мы далее называем "когнитивной". Основным носителем когнитивной среды - мозг познающего индивида, который может расширять ее за счет внешних средств. Наиболее мощным из них в настоящий период является компьютер. Поэтому вполне логично ввести понятие компьютерной когнитивной среды, в которой

протекает часть процессов, связанных с познанием избранного объекта, и которая должна быть построена так, чтобы усиливать информационные и мыслительные возможности исследователя.

Используя компьютерную метафору к познавательной деятельности субъекта, когнитивные процессы можно разделить на три вида: хранение информации, переработка информации и передача информации. В соответствии с этим части когнитивной среды, где протекают перечисленные процессы будем называть информационной, операционной и коммуникационной.

Стремясь к тому, чтобы компьютерная когнитивная среда была как можно более естественным продолжением когнитивной среды исследователя будем представлять архитектуру исторических информационных систем как совокупность трех взаимосвязанных сред: информационной, операционной и коммуникационной.

Информационная среда обязана обеспечить когнитивно обусловленные структуры для хранения информации как непосредственно относящейся к историческому исследованию, так и связанной с функционированием когнитивной среды как информационной системы.

Историко-исследовательской информации кроме большого объема присуща также значительная разнородность. В зависимости от характера источников и поставленных исследовательских задач историческая информация может быть в большей или меньшей степени структурирована и представлена либо в одной из однородных форм (текстовой, графической, аудиальной, визуальной), либо путем их сочетания. Причем хорошо структурированная информация может находиться на различных уровнях семантичности - от "данных" до "знаний".

В настоящее время для представления и обработки информации в компьютерной среде разработано множество информационных технологий. Своя технология соответствует каждой из перечисленных однородных форм представления информации. Хорошо структурированная информация, как правило, представляется (исходя из уровня требуемой семантичности) с помощью либо технологии баз данных, либо технологии знаний, а для лучшей организации слабо структурированной информации рекомендуется использовать гипертекстовые технологии, в том числе (для сложных форм представления информации) технологии гипер- и мультисред. В общем случае очевидна необходимость сочетания возможностей всех упомянутых информационных технологий. При этом в качестве интегрирующей можно использовать технологию мультисред, которая позволяет объединить под одной крышей информацию, различную по структурированности, уровню семантичности и форме представления.

В структуре операционной среды можно выделить следующие основные группы средств:

- средства поддержки информационной среды, обеспечивающие интегрированное представление разнородной историко-исследовательской

информации или (для однородной информации) по одной из указанных выше технологий; - проблемно-ориентированные средства, реализующие компьютеризированные методы решения историографических задач, основанные на обработке данных или базирующиеся на обработке знаний;

- средства поддержки коммуникационной среды с преобладанием одного из способов отображения информации: вербального или образного, либо равноправно обеспечивающие все виды внешнего представления коммуникационной информации;

- управляющие средства, которые обеспечивают координацию функционирования системы в целом и отдельных ее сфер - поддержки информационной среды, средств решения исследовательских задач, поддержки коммуникационной среды, сервисных средств;

- сервисные средства, предназначенные для обслуживания информационной и коммуникационной сред, проблемно-ориентированных средств и сферы управления.

К коммуникационной среде относятся средства, обеспечивающие взаимодействие пользователей с системой, передачу информации внутри системы и в когнитивной среде. Средства общения пользователя с системой могут либо сочетать вербальный и образный способы взаимодействия, либо преимущественно использовать один из них.

Предложенная типовая архитектура ИИС как бы устанавливает самые общие, родовые черты, которыми в необходимой мере должна обладать любая система данного класса. Примером системы, принадлежащей к классу ИИС, может служить описываемая далее система "Просис", которая разработана специально для компьютерной поддержки проекта "Российские историки XVIII - начала XX вв." Разновидность (подкласс) ИИС, выступающих в качестве компьютерной когнитивной среды для научно-биографических исследований будем называть "просопографические информационные системы" (ПИС). Очевидно, что информационная система "Просис", будучи исторической, является также и просопографической.

Просопографические информационные системы: принципы построения и реализации

Принципы построения и реализации просопографических информационных систем мы рассмотрим на примере текущей версии системы "Просис". Названная система базируется на современных информационных технологиях, в частности, технологии баз данных и гипертекстовой технологии, и построена в соответствии с типовой архитектурой ИИС.

В структурах информационной среды системы "Просис" может храниться представленная в текстовой форме как сильно, так и слабо структурированная информация. Основой организации хранения сильно структурированной информации является технология баз данных, а слабо структурированной - гипертекстовая технология, которая также служит базой для организации интегрированного хранения всей относящейся к проекту информации.

В части информационной среды, организованной как реляционная база данных, может содержаться следующая информация: сведения об историках, представленные в форме разделов анкет, со ссылками на источники получения данных и составителей и примечаниями, если они имеются, к каждому из разделов; библиографические данные об источниках; сведения о составителях; совокупность словарей-кодов, которые представляют собой мнемонически значимые сокращения возможных значений реквизитов анкеты с их полной расшифровкой; общесистемная информация: список анкет, их количество и картина заполненности, количество источников, список словарей, список составителей и т.д.

Гипертекстовая часть информационной среды предназначена для организованного хранения всей остальной, связанной с проектом информации, которая не укладывается в жесткие структуры реляционной базы данных. К такой информации относятся описания исследуемых вопросов и решаемых задач, различного рода выписки из источников и литературы (например уникальные сведения об историках), мысли исследователей, аннотации источников, научных работ и методов исследования, формулировки запросов к базе данных, фрагменты текстов подготавливаемых публикаций и т.д. Как известно, главное достоинство гипертекстовой технологии в том, что она позволяет связать самым произвольным образом самые разнообразные информационные объекты. Поэтому в гипертекст системы "Просис" могут включаться не только тексты, но и фрагменты базы данных: выборочные анкеты или библиографические сведения об отдельных источниках. Этим самым гипертекст "Просис" обеспечивает интеграцию слабо структурированной информации с сильно структурированной и дополнительные возможности связывания информационных единиц базы данных.

В качестве узлов гипертекстовой сети выступают так называемые "объекты гипертекста" (ОГТ), которые играют роль представителей информационных объектов (ИО) в гипертексте. Каждый ОГТ имеет уникальный номер и представляет собой идентификатор (поисковый образ) ИО, код типа ИО (текст, анкета, источник или др.) и указатель на ИО. С одним ИО может быть связано несколько ОГТ, представляющих его с разных сторон. При такой организации сам ИО не считается компонентом гипертекста и существует независимо от него. Исходя из объективного характера ИО или своих субъективных соображений, исследователь имеет возможность объединять ОГТ в группы, именуемые "ассоциациями". Один и

тот же ОГТ может входить в несколько ассоциаций. Все связи гипертекста "Просис" двунаправленные и определяются не только соединяемыми ОГТ, но и тем, в рамках каких ассоциаций эти ОГТ связываются. В зависимости от отношения к ассоциациям связи делятся на внутриассоциативные и межассоциативные.

Операционная среда системы состоит из относительно самостоятельных, но связанных подсистем: "Анкеты"; "Источники"; "Составители"; "Словари"; "Гипертекст".

В подсистеме "Анкеты", основной в системе "Просис", предусмотрено четыре режима работы: "Ввод", "Редактирование", "Просмотр" и "Исследование". В режиме "Ввод" осуществляется первоначальное наполнение анкет данными. С режимом "Ввод" тесно связан режим "Редактирование", который обеспечивает корректировку уже введенной информации, дозаполнение разделов анкет или удаление данных из некоторых пунктов. В обоих режимах там, где это возможно, предусмотрен контроль вводимых данных, повышающий степень целостности информационной среды системы. Режим "Просмотр", тяготеющий по своему целевому смыслу к режиму "Исследование", предназначен для просмотра содержания анкет. Просмотр может выполняться как по всей совокупности заполненных анкет, так и по анкетам, отобранным в соответствии с определенными условиями, при этом предусмотрено два варианта доступа к разделам анкет: анкета-разделы, раздел-анкеты.

Набор и содержание процедур, предоставляемых пользователю режимом "Исследование", определяется задачами исследования конкретной проблемы и возможностями их компьютерного решения. В текущей версии системы "Просис" предусматривается использование лишь методов, базирующихся на обработке данных. В перспективе, после проведения специальных исследований, предполагается подключение и средств, основанных на обработке знаний.

Подсистема "Источники" предназначена для введения сведений об источниковой базе исторических исследований. В ней имеется три режима работы: "Ввод", "Редактирование" и "Просмотр", функции которых в основном понятны из их названий. Аналогичными возможностями обладает и подсистема "Составители", которая схожа с подсистемой "Анкеты".

Поддержка словарей-кодов системы осуществляется подсистемой "Словари", в основе которой лежит концепция "базовый словарь - рабочий словарь". В базовых словарях хранятся все известные на данный момент значения реквизита, а в рабочих - лишь необходимые для текущей работы. В соответствии с этим подсистема работает в двух режимах: "Базовые словари" и "Рабочие словари". В первом режиме выполняется дополнение, редактирование и просмотр базовых словарей, а во втором - формирование рабочих словарей на основе базовых, их упорядочение и при необходимости просмотр.

Для интеграции всей относящейся к проекту информации и работы со слабо структурированными данными служит подсистема "Гипертекст". Она функционирует в трех режимах: "Формирование", "Редактирование" и "Просмотр". В режиме "Формирование" выполняются операции по созданию ОГТ и ИО, корректировке идентификаторов ИО, уничтожению ОГТ, включению ОГТ в ассоциации и исключению из них или переводу из одной ассоциации в другую, установлению и устранению связей между ОГТ. Всем операциям, за исключением операции создания, предшествует операция поиска для выделения тех объектов гипертекста, над которыми должна быть выполнена требуемая операция. Поиск может осуществляться любым из следующих способов: в соответствии с номерами ОГТ или ИО, по заданным признакам идентификаторов или путем прохождения по связям. Он может происходить по всему гипертексту, в пределах избранного типа ИО или внутри заданной ассоциации. Любая произведенная в режиме "Формирование" операция, кроме поиска, изменяет состояние гипертекста. Режимы "Редактирование" и "Просмотр", различаясь, обладают общими возможностями, в том смысле, что оба позволяют находить интересующие пользователя ОГТ и ИО и просматривать их. Но при этом в режиме "Редактирование" возможно еще и изменение содержания ИО без нарушения состояния самого гипертекста.

Учитывая особенности историков как конечных пользователей компьютеризированных систем, в основу построения коммуникационной среды системы "Просис" положен принцип максимально возможного приближения форм их общения с компьютером к профессионально привычным. Поэтому сведения об историках представляются в виде традиционных бланков разделов анкет с модификацией в сторону повышения эффективности их восприятия, информация об источниках отображается в форме карточки, данные о составителях представляются аналогично сведениям об историках, а вид словарей-кодов мало чем отличается от принятого. Представление на экране гипертекста основывается на гипотезе, что исследователя в данный момент чаще всего интересует содержание лишь одного из ИО и его связи с другими.

Управление работой системы "Просис", т.е. выбор необходимых режимов работы или действий, осуществляется исключительно методом "меню" и в терминологии, естественной для пользователя-историка.

Взаимодействие со средствами решения исследовательских задач происходит по типу: "вопрос системы - ответ пользователя" с подсказкой возможных ответов. При этом выходная информация, являющаяся конечным результатом компьютерных исследований, выводится в основном в вербальной форме с графическим дополнением, усиливающим ее восприятие. Во всех способах взаимодействия соблюдаются принципы оптимизации информативности экранного кадра по отношению к количеству затраченных на это знаков и рационального сочетания естественности интерфейса для

пользователя-историка с преимуществами в работе с информацией, предоставляемыми компьютерной средой.

При разработке и реализации системы мы делали акцент на обеспечении комфортности информационной деятельности историка, не забывая при этом об эффективном использовании компьютерных ресурсов. Выбор базовых технических и программных средств определялся их доступностью для историков и опробованностью. Таковыми, по нашему мнению, являются персональные компьютеры типа IBM PC и программные системы: ОС MS DOS, редакторы текстов Lexicon или Word, СУБД FoxBase Plus или Clipper. Представленный вашему вниманию вариант системы "Просис" в настоящее время находится в опытной эксплуатации, которая в аспекте наполнения ее базы данных носит реальный характер.

Рекомендации по осуществлению проекта с помощью системы "Просис"

В процессе реализации проекта "Российские историки XVIII - начала XX вв." с использованием системы "Просис" можно выделить следующие фазы:

- сбор информации об историках;
- подготовка системы; ввод исходной информации;
- редактирование исходной информации;
- постановка исследовательских задач;
- выбор или разработка методов решения задач;
- решение историко-исследовательских задач.

В рамках настоящей статьи мы дадим ряд наиболее важных рекомендаций, касающихся лишь первых четырех фаз. Рассмотрение вопросов реализации проекта, связанных с остальными фазами, из-за их особого характера оставляем за ее пределами.

Ввиду громадного объема материала сбор информации разумно вести по частям и в определенном порядке. Среди множества возможных декомпозиций, по нашему мнению, наиболее рационально деление персоналий прежде всего по периодам, затем внутри их по регионам, а внутри регионов по контингентам. Что касается порядка накопления сведений в рамках очерченной группы, то собирать данные лучше всего в такой последовательности: элита; известные, известные но малоизученные историки; неизвестные историки. Такой порядок наиболее эффективен с точки зрения поиска источников и верификации применяемых методов решения исследовательских задач.

Работа по сбору информации может осуществляться как индивидуально, так и коллективно. Из ряда вариантов организации групп "сборщиков" заслуживающими внимания нам представляются следующие:

- "координатор - коллеги": для историко-исследовательских учреждений;

- "учитель - ученики": для вузов;
- "центр - регионы": способ координируемого сбора информации коллективами России и ближнего зарубежья.

В фазе подготовки, исходя из целей и задач исторического исследования, производится настройка среды системы и определяется конфигурация ее операционных и коммуникационных средств. Настройка информационной среды заключается в приведении всех файлов системы в исходное состояние, установке паролей доступа, заполнении базовых словарей, формировании рабочих словарей и их упорядочении. Генерация конфигурации средств, обеспечивающих операционно-коммуникационную среду, в основном состоит в комплектовании соответствующей совокупности программных модулей и их настройке.

Из всех перечисленных фаз наиболее ответственными являются фазы ввода и редактирования исходной по отношению к историческому исследованию информации. Для предотвращения искажения вводимых данных необходимо минимизировать их путь к системе. В идеале сведения, извлекаемые из источника, должны вводиться непосредственно в компьютер, минуя какие-либо промежуточные носители. Надлежащим образом следует продумать и реализовать организационную сторону заполнения информационной базы системы.

Ввод исходной информации предпочтительнее начинать со сведений о составителях анкет, затем следует ввести библиографические данные об источниках, которые послужили основой для составления анкет, и только после этого можно заполнять разделы анкет со ссылками на источники и составителей.

Редактирование введенных сведений может и должно выполняться многократно и тщательно. Только так можно обеспечить достоверность, полноту и непротиворечивость информации по исследуемой исторической проблеме. Действия по редактированию рекомендуется выполнять в следующем порядке:

- просмотр ранее введенных данных с исправлением ошибок и, при необходимости, удалением некоторых из значений реквизитов;
- просмотр существующих информационных объектов и дополнение новыми данными;
- контрольный просмотр;
- запись измененных данных в память системы.

Очевидно, что нарушение приведенных рекомендаций по редактированию уменьшает гарантию качества информационной базы для исторических исследований.

Заключение

Даже самые совершенные возможности системы "Просис" и большая комфортность общения с ней не гарантируют успеха в осуществлении

изложенного выше проекта, если не будут собраны как можно более полные и качественные сведения о персоналиях отечественной исторической науки. Поэтому публикуя эту статью мы стремимся не только обнародовать результаты своей работы, но и надеемся привлечь к сбору материала и проведению соответствующих исследований всех, кто заинтересован в сохранении и реконструкции портрета строителей российской истории.

ПРИМЕЧАНИЯ

1. Интервью с д-ром Леном Брере, генеральным секретарем Международной Ассоциации "History and Computing" (АНС) // Информационный бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории РАН и ассоциации "История и компьютер", N 8, март, 1993,- с.98.

2. Талер М. Что такое "источнико-ориентированная обработка данных"; что такое "историческая информатика"? // История и компьютер: Новые информационные технологии в исторических исследованиях и образовании. St.Katharinen, 1993, с.5.

3. Интервью с д-ром Леном Брере... Ук. Соч., с.99-100.; Бородкин Л.И. Историческая информатика: Горизонты 90-х годов // Методологические проблемы исторической науки. Сб. статей по материалам международной научной конференции (Минск, 29 сентября - 1 октября 1992 г.). Минск, 1993, С.88.

4. Найсер У. Познание и реальность. М.,1981, с.21.

II. АРХИВЫ МАШИНОЧИТАЕМЫХ ДАННЫХ

АРХИВЫ МАШИНОЧИТАЕМЫХ ДАННЫХ И ПРОБЛЕМЫ АРХИВИРОВАНИЯ МАШИНОЧИТАЕМЫХ ИСТОРИЧЕСКИХ ИСТОЧНИКОВ

И.М.Гарскова

Для нынешнего этапа развития науки характерен все возрастающий поток информации в машиночитаемой форме. С одной стороны, в специализированные научные архивы поступают коллекции данных, содержащие результаты конкретных исследовательских проектов, с другой стороны - в традиционные государственные архивы сдается множество документов текущего учета, часто это большие, систематически обновляемые и пополняемые массивы (например, текущая документация государственных учреждений или периодические переписи). Однако и государственные, и специализированные архивы сталкиваются с похожими проблемами, при попытке решения которых сказывается недостаток опыта работы с "электронными записями" /1/.

Опыт "старых" архивов машиночитаемых данных (подобных Межуниверситетскому Консоциуму по политическим и социальным исследованиям в США /2/) не всегда на пользу новым, если он не осмыслен критически. Дело в том, что радикальные изменения в технологии хранения данных привели к изменениям принципов хранения и распространения электронных записей. Так, острота дискуссий относительно стандартов хранения информации (агрегирования, стандартизации кодировок или форматов записи файлов) значительно снизилась вместе с расширением возможностей машинной памяти, специализированного программного обеспечения для создания и модификации баз данных. При обсуждении принципов хранения машиночитаемых версий исторических источников отмечается тенденция отказа от кодирования и трансформирования исходной информации в пользу максимально бережного сохранения как структуры, так и текста первичного источника: исследователь, имеющий возможность выполнить такую работу (например, при определенной поддержке соответствующего

архива), параллельно может создавать и производные файлы, содержание и структура которых подчинены задачам конкретного исследования. Таким образом, обеспечиваются интересы не только исследователя, но и архива, заинтересованного в создании максимально близкой оригиналу машиночитаемой копии источника. Меняются и стандарты распространения машиночитаемых данных: от пересылки полных массивов на магнитных лентах и сопутствующей бумажной документации по почте до передачи нужных выборок данных, документации и необходимого программного обеспечения непосредственно на компьютер пользователя /3/.

Наконец, возникают качественно новые проблемы при архивировании и документировании нетрадиционных данных, не похожих на привычные прямоугольные таблицы или плоские файлы, (например, реляционные базы данных со сложными связями или файлы, включающие изображения или гипертекст, в котором данные неотделимы от программного продукта и т.п.) /4/.

Поэтому сейчас как в научных кругах, так и в руководстве государственных архивных служб наступила пора переосмысления принципов деятельности таких архивов и выработки единой политики в области архивирования машиночитаемых данных /5/. Интересы исследователей и архивистов в этом направлении в основном совпадают; в последнее время активизировалось и международное сотрудничество, поскольку в разных странах деятельность как традиционных, так и электронных архивов регулируется хотя и разными, но во многом сходными законами.

1. Основные проблемы функционирования архивов машиночитаемых данных

В чем же состоят основные проблемы деятельности архивов машиночитаемых данных? До сих пор ни в одной стране нет единой государственной политики по машиночитаемым файлам данных, которые хранятся в отдельных центрах без особых перспектив на будущее. Нужны четкие принципы стандартов хранения данных и защиты информации, в частности, требуют обсуждения критерии хранения и уничтожения информации, вопросы ее защиты от уничтожения или несанкционированного копирования (можно упомянуть и такой аспект защиты информации как специальные правила об уничтожении записей о персоналиях, имеющих "порочащий" их характер, после того, как их актуальность для государства утрачивается). Не решены вопросы "длительности жизни" машиночитаемых файлов в зависимости от носителей, выбора наиболее адекватного аппаратного и программного обеспечения, причем при переводе информации на новое программное или аппаратное обеспечение не всегда ясно, следует ли продолжать хранить старые версии файлов /6/.

Таким образом, наиболее важными проблемами в деятельности архивов машиночитаемых данных являются:

- 1) разработка принципов хранения информации;
- 2) организация защиты информации;
- 3) активизация вторичного использования информации;
- 4) стандартизация критериев описания данных;
- 5) оценка стоимости создания и архивирования машиночитаемых файлов;
- 6) регулирование отношений между архивом и разработчиками машиночитаемых данных /7/.

Рассмотрим некоторые из них более подробно.

Регулирование отношений между архивом и разработчиком данных

Как правило, разработчик заполняет несколько специальных форм, фиксирующих назначение, форматы и содержание файлов, средства ввода/вывода, а также список всех переменных с их именами и кодами. В задачу архива входит проверка соответствия файлов их описаниям, затем архив присваивает массиву данных соответствующие реквизиты, создает резервные копии, оценивает качество документации с точки зрения ее пригодности для распространения. Документы, которые подписываются автором и архивом, оговаривают, в частности, условия доступа к данным /8/.

Существуют также специальные правила, регулирующие отношения между архивом и вторичными пользователями. Потенциальные пользователи также подписывают с архивом документы, включая обязательство корректного цитирования источника получения данных в каждой своей публикации, связанной с ними. Архив обязуется предоставить все необходимые данные вместе с документацией на определенном носителе /9/.

Все эти проблемы легче решать, учитывая, что процесс архивирования во многом похож на публикаторскую деятельность, в которую традиционно включаются редактирование данных и соблюдение авторского права /10/. В аспекте авторского права должны решаться вопросы о праве на массивы, созданные с использованием опубликованных, т.е. общедоступных данных, или вопросы о праве на данные, созданные вторичным пользователем с использованием массива, который он получил из архива (и право на который сохраняет его разработчик).

Из перечисленного становится ясно, какое важное место в деятельности архивов занимают вопросы авторского права. Право разработчика на машиночитаемые данные должно, как и право на другие виды интеллектуальной собственности, регулироваться специальным законодательством.

В России осенью 1992 г. был принят, а с 1 января 1994 г. вступил в полную силу закон "О правовой охране программ для электронных вычислительных машин и баз данных", регулирующий отношения, связанные с созданием, правовой охраной и использованием программ для ЭВМ и баз

данных. Этот закон предоставляет правовую охрану базам данных как выпущенным, так и не выпущенным в свет и представленным в объективной форме, на том же уровне как публикуемым сборникам.

Предполагается, если не доказано обратное, что базы данных есть результат творческой деятельности их авторов; следствием этого является положение об охране прав на базы данных независимо от того, являются ли данные, на которых они основаны или которые они включают, объектами авторского права, однако авторское право на базу данных признается законом при условии соблюдения авторского права на все ее составляющие. Интересно, что авторское право на базу данных не препятствует другим лицам осуществлять создание собственных баз данных с использованием тех же первичных материалов.

Для возникновения авторского права достаточно уже самого факта создания базы данных и не требуется соблюдения никаких формальностей, хотя можно зарегистрировать ее в установленном порядке. Автором базы данных признаются только физические лица, которые получают на нее соответствующие личные и имущественные права. Если база данных создана в порядке выполнения служебных обязанностей, то имущественные права на нее могут принадлежать работодателю. Имущественные права, включающие право выпуска в свет, воспроизведения, распространения, модификации или иного использования базы данных, могут передаваться полностью или частично другим физическим или юридическим лицам по письменному договору. Договором же определяется использование базы данных третьими лицами.

Принципы хранения информации в архивах машиночитаемых данных

Поступающие в архив файлы имеют, как правило, разные форматы; данные могут быть: а) исходными; б) преобразованными в содержательном и техническом плане (например, агрегированными); в) преобразованными и кодированными; г) в специальных форматах; д) в виде программы с включенными данными /11/.

Как правило, архив предпочитает т.н. ASCII-файлы в виде обычных прямоугольных таблиц (плоские файлы), т.к. этот формат является максимально независимым как от программного, так и от аппаратного обеспечения. Если структура данных более сложная, то удобнее провести декомпозицию, разложив ее в плоские файлы в обычном формате. Исходная структура может быть при этом воссоздана из "элементарных" файлов с помощью специальных программ. Например, это относится к базам данных, доля которых в архивах постоянно возрастает. Формат хранения базы данных зависит от использованного программного обеспечения, и архив принимает решение либо о сохранении этого формата, либо о модификации файлов в более простые.

Хотя эта проблема решается довольно просто для баз данных, она представляет принципиальные сложности для таких структур как гипертекст (увязанная в единую сеть разнородная информация – тексты, графики, базы данных, электронные таблицы и т.п.), просто не существующий отдельно от программного обеспечения. Тем не менее, возможно в идеале представить, что данные и документация в машиночитаемой форме будут поставляться пользователю на диске вместе с соответствующим программным обеспечением, дающим возможность реконструкции сложных структур из нескольких независимых файлов /12/.

Стратегия работы архива машиночитаемых данных с традиционными источниками

Пока превращение бумажных источников в машиночитаемый вид не является центральной задачей таких архивов, однако развитие методов сканирования и оптического распознавания дает и такую возможность /13/.

Если ставится задача перевода в электронный формат традиционных источников, можно назвать три способа такого перевода: 1) полный перевод (наиболее трудный и дорогостоящий способ, применимый лишь для наиболее важных массивов); 2) сканирование и последующее распознавание; 3) сканирование и хранение образа данных и документации (более быстрый способ, но создающий файлы большого объема и в отличие от первых двух не позволяющий проводить информационный поиск) /14/.

Документирование машиночитаемых данных

Одной из основных трудностей в работе архива является то, что разработчики данных крайне редко заботятся о нуждах вторичных пользователей (т.е. в первую очередь об адекватной документации). Дело осложняется тем, что файлы, зависящие от программного обеспечения, создают дополнительную преграду между ними и их потенциальными пользователями. Это относится не только к исследовательским файлам – в 80-е гг. сокращается и объем описания машиночитаемых данных, поступающих из официальных источников, ввиду сокращения расходов на оформление. Архивы также не имеют средств для улучшения качества документации, поэтому пока наиболее надежным способом получения документации является непосредственный контакт с разработчиком /15/.

По своему назначению можно выделить три уровня документации: пользовательский, описывающий цель, источник, организацию, содержание и характеристики машиночитаемых файлов для возможного вторичного использования; административный, включающий официальные документы, регулирующие отношения архива с пользователями (см. предыдущий раздел)

и, наконец, каталожный, предназначенный для автоматизированного поиска в архиве нужных файлов /16/.

Идеальной документацией можно считать полную информацию о процедуре исследования, принципах создания и обработки данных, включая листы кодировки (codebooks), и информацию о форматах всех файлов /17/. Пользовательская документация должна включать информационную часть (название, авторство, аннотацию, ключевые слова, описание программного и аппаратного обеспечения, краткое описание методики создания массива); "историю создания" исходного источника (принципы измерения, географические и хронологические рамки, оценку качества данных, процедуры коррекции ошибок); технические характеристики файлов (СУБД, DOS, число записей, количество и длины полей, защита информации); словарь данных; одномерные распределения некоторых важных показателей; дополнительную информацию (некорректности в данных, версии файлов и т.п.) /18/. Формат документации должен быть независимым от программного обеспечения.

2. Специфика документирования машиночитаемых исторических данных

К концу 1980-х - началу 1990-х гг. в развитии архивов машиночитаемых данных начинается новый этап: планирование и создание в их структуре исторических отделов (Германия, Дания, Англия), появились и специальные исторические архивы - Национальный Архив машиночитаемых исторических данных (NHDA) в Голландии. В первую очередь это было связано с быстрым и довольно неупорядоченным ростом числа машиночитаемых коллекций данных в исторических исследованиях. Внимание к источнику, ориентация на специфику исторических источников объясняется тем, что при всей близости истории и социальных наук, на которые ориентировались традиционные архивы машиночитаемых данных, источники в историческом исследовании играют совершенно особую роль.

Историк находится в ином положении по отношению к данным, чем, например, социолог, не говоря уже о представителях естественных наук. Социолог формирует информационный массив, проводя опрос или обследование, историк же вынужден искать (и не всегда успешно) интересующие его сведения в уже имеющихся источниках, иногда лишь косвенно связанных с поставленной проблемой. По образному выражению Р. Морриса, "у историка уже есть ответы, и он должен правильно понять, как поставить к ним вопросы" /19/.

В чем состоит специфика машиночитаемых исторических источников с точки зрения их содержания и структуры? Помимо того, что историки гораздо чаще, чем социологи, работают с нарративными источниками более сложной структуры, чем анкетный материал, типичный для социальных

наук, можно указать и на другие важные отличия. Например, приходится учитывать то обстоятельство, что исторический источник зачастую анализируется в ином аспекте и с иной целью, не совпадающей с целью его создателя (чем и диктуется иногда трансформация исходных данных), или укажем на проблему неполноты данных (утраты, пропуски, ошибки в источнике). Далее, характер многих источников таков, что исследователь использует лишь часть информации (например, выбирает некоторые фактические данные из обширного текстового источника). Все эти особенности трансформации (включая кодирование, агрегирование и т.п.) должны быть, очевидно, тщательно обоснованы. Важно подчеркнуть, что историки, возможно потому, что они позже включились в процесс компьютеризации, предпочитают работать с более сложными по структуре файлами, чем прямоугольные числовые таблицы (т.е. с базами данных, свободным текстом и др.) и в связи с этим используют более сложное программное обеспечение.

Уже со второй половины 1980-х гг. историки, работающие в области исторической информатики в разных странах, начали систематически обсуждать актуальные проблемы и принципы компьютерного архивирования и распространения машиночитаемых данных. Большую роль в этой деятельности играет Международная Ассоциация "History and Computing" (АНС), проявившая интерес к этим проблемам практически со дня своего создания (1986 г.). На трех конференциях АНС - в Геттингене (1985 г.), Граце (1986 г.) и Париже (1987 г.) - стоял вопрос о стандартизации описания машиночитаемых исторических данных. На конференции в 1989 г. Центром QUANTUM и АНС была предпринята попытка создания единого каталога машиночитаемых данных в области историко-социальных исследований в европейских и американских архивах; большую роль сыграла конференция "Проблемы архивирования и распространения исторических машиночитаемых данных" (Лейден, 1990 г.), результаты которой опубликованы в специальном выпуске бюллетеня Голландской национальной ассоциации "History and Computing" "Data, Computers and the Past". Отметим также последние европейские семинары по проблемам документирования машиночитаемых исторических источников (май 1993 г., Оденсе, Дания) и проблемам вторичного использования в исторических исследованиях электронных информационных ресурсов (июнь 1993 г., Лондон). Начиная с 1989 г. в каждом номере журнала HSR (Historical Social Research) публикуется информация о базах и коллекциях машиночитаемых данных в области историкосоциальных исследований, имеющих в различных национальных архивах. Большое внимание к этой проблематике проявляют и другие журналы ("Computers and the Humanities" и "History and Computing").

На конференции АНС в Геттингене, посвященной созданию, координации и использованию универсальных междисциплинарных банков исторической информации, была создана рабочая группа для разработки стандарта

описания исторических машиночитаемых данных ("Троица", по выражению Х.-Ю. Маркера), в которую вошли Х.-Ю. Маркер, К. Шурер (Кембриджская группа по изучению истории населения и социальных процессов, Кембридж, Англия) и Г.Рейнке (Quantum - Центр историко-социальных исследований, Кельн, Германия). Предложения, подготовленные этой рабочей группой, были сформулированы к 1988 г. Они базировались на схеме SSD (Standard Study Description) - стандарте описания социологического обследования (см. приложение 1), адаптированном для учета специфики исторических источников /20/. Параллельно в Голландии П. Доорн и его коллеги из Национального Архива машиночитаемых исторических данных разрабатывали детальную схему описания машиночитаемых исторических источников, которая отличалась от схемы, предложенной "Троицей" Рейнке-Шурер-Маркер, в основном тем, что за единицу описания брался не исходный источник, являющийся результатом исследования (study), а коллекция машиночитаемых файлов (dataset). Тем не менее, оба подхода имеют больше сходства, чем различия.

Возможно, первый больше соответствует точке зрения архивиста: он стандартизирует описания, а также структуру файлов и листов кодировки (codebooks), что важно для облегчения обмена информацией, издания периодических каталогов и функционирования электронного каталога. Особенностью этого подхода к описанию данных является то, что исследование, базирующееся на нескольких источниках, требует создания соответствующего количества описаний по каждому из них, равно как и описания "сводного" файла с указанием его связей с каждым из первичных файлов по отдельным источникам. Несмотря на некоторую громоздкость такого подхода, он дает возможность вторичному пользователю без труда обращаться к первичным файлам, если его не интересуют вторичные, несущие на себе влияние конкретного исследования.

Отличительной чертой стандарта описания данных исторических источников, принятого во многих европейских архивах, является внимание не только к машиночитаемым данным социологических обследований, уже хранящихся или поступающих на хранение в архив. Так, стандарт, предложенный Датским архивом (DDA), пополнен несколькими специальными разделами, связанными с подробным и тщательным описанием процедур перехода "от источника к данным" /21/. Эти разделы включают сведения о возможности доступа к исходным данным, о целях их создания и о процедурах трансформации источника при создании его машиночитаемой версии (см. схему, предлагаемую DDA, и анкету для разработчика, служащую основой для заполнения этой схемы, в приложении 2).

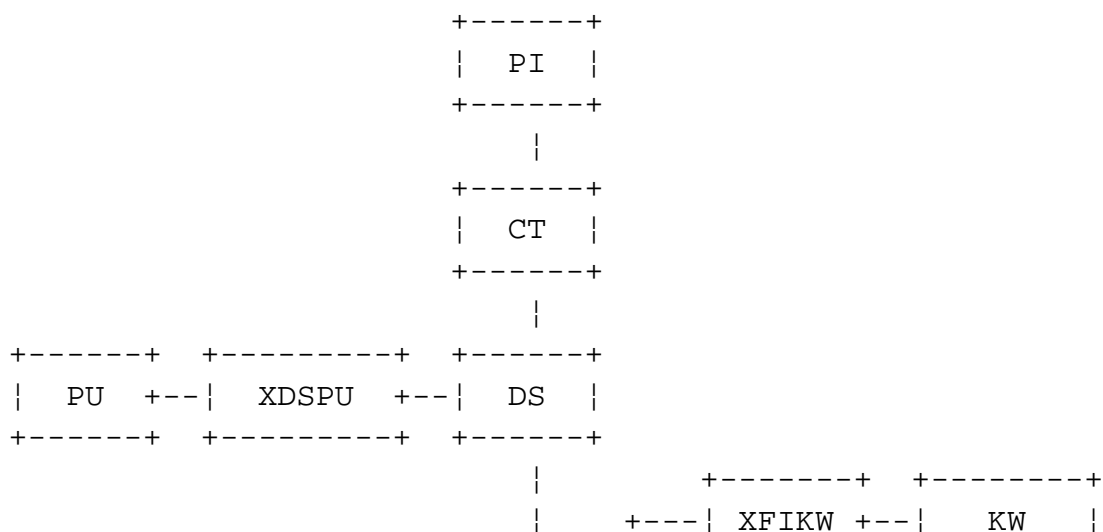
Второй подход проводит Голландский Архив исторических данных (NHDA), где разработана собственная схема описания исторических (машиночитаемых) данных (HDDS - Historical Datasets Description Scheme) /22/. Как уже сказано, за основную единицу описания здесь берется коллекция данных (и отсюда название единицы описания -

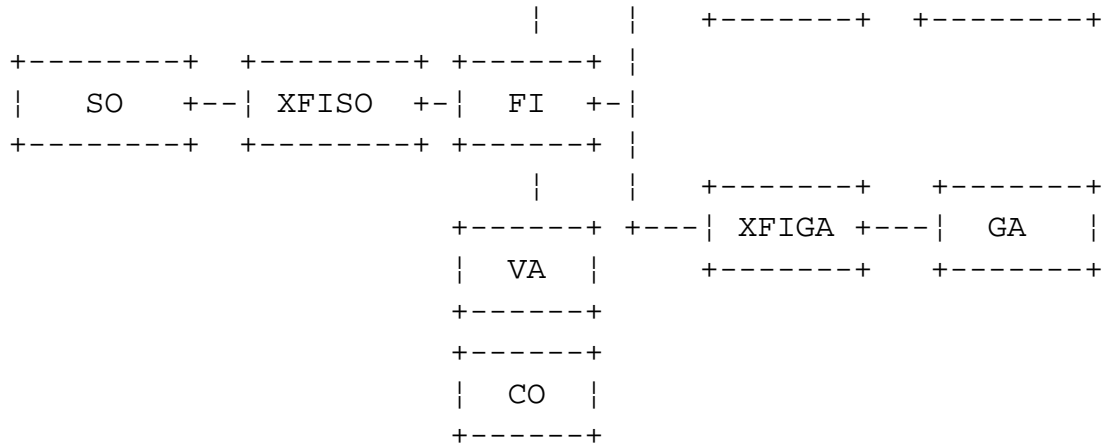
dataset), т.к. большинство исследований базируется на нескольких источниках и многие источники используются многими исследователями. Безусловно, это позиция, типичная не для "чистых" архивистов, а для историков-исследователей, создающих в процессе своей работы новые, ранее не существовавшие (например, сводные из нескольких) источники. Более того, данные первичных источников часто кодируются, агрегируются или трансформируются каким-либо иным образом, т.е. файлы, с которыми работают исследователи, никоим образом не являются простыми копиями исходных или "бумажных" источников, но представляют собой самостоятельные "производные" источники.

Точка зрения NHDA также требует очень аккуратного описания как исходных источников, так и методов перехода "от источника к данным". В этом отношении обе позиции можно оценить как довольно близкие и допускающие окончательное формирование единого европейского стандарта документирования исторических машиночитаемых данных. Более того, HDDS предусматривает специальный интерфейс с SDD с целью поддержания возможностей обмена данными между вновь создаваемыми архивами исторических данных и уже давно существующими архивами данных социальных исследований (схема HSSD приводится в приложении 3).

В NHDA разработана специальная модель описания, имеющая структуру реляционной базы данных. Эта модель предполагает, например, что каждый источник описывается только один раз, но может быть связан с описанием любого исследования, использующего этот источник; то же самое относится и к описанию каждой публикации, которая связывается с любым использованным ее автором источником и всеми исследовательскими проектами, имеющими на нее библиографические ссылки и т.д..

Модель, о которой идет речь, т.е. схема реляционных связей между отдельными элементами HDDS (коллекция данных, файл, переменная, код переменной, публикация, исследователь или организация) может быть представлена в следующем виде /23/:





Обозначения элементов схемы:

- DS - коллекция данных; CT - контакты;
- PI - исследователь или организация; FI - файл;
- XFISO - связь файл-источник; SO - источник;
- XFIKW - связь файл-ключевое слово; KW - ключевое слово;
- XFIGA - связь файл-территория; GA - территория;
- XDSPU - связь колл.данных-публикация; PU - публикация;
- VA - переменная; CO - код;

Рис. 1. Реляционная модель описания данных

Очевидно, что при рутинном способе документирования такая сложная структура описания потребовала бы от исследователя значительных затрат времени и сил на ее изучение и заполнение большого числа разных анкет. Однако в NHDA создана специальная программа "Сделай сам" (DIY - Do-It-Yourself) автоматического документирования и ведения каталогизации /24/. Задавая разработчику коллекции данных простые вопросы о характеристиках этих данных, программа позволяет быстро и эффективно создать машиночитаемую документацию и при необходимости напечатать ее в кратком или полном варианте. Безусловно полезным, например, и с этим согласились все участники семинара, при любой научной публикации, даже не связанной с передачей данных в архив, является приложение с кратким описанием использованных данных /25/.

В деятельности организуемого в России Архива машиночитаемых исторических данных направление, связанное со стандартизацией и документированием, считается одним из наиболее важных. С одной стороны, Архив поддерживает контакты с организациями, занимающимися ведением электронного всероссийского каталога машиночитаемых данных, публикацией информации о таких данных и разработкой стандартов их документирования. До недавнего времени такой организацией являлся ВИМИ

- Всесоюзный (Всероссийский) Институт межотраслевой информации, который в 1990 г. издал первый выпуск каталога баз данных /26/, содержащего описание 1400 баз по всем отраслям науки, техники и народного хозяйства. С 1992 г. функции всероссийского информационного центра перешли к НТЦ "Информрегистр" Комитета Российской Федерации по информатизации. В опубликованных этим центром втором и третьем выпусках каталога уже появились сведения о базах исторических машиночитаемых данных /27/.

Другим направлением деятельности Архива является разработка российского стандарта описания машиночитаемых исторических источников, первая версия которого была опубликована в Информационном бюллетене Комиссии по применению математических методов и ЭВМ в исторических исследованиях N 2 в 1991 г. /28/ и по этому стандарту документировались те базы данных, описания которых можно найти в последующих номерах бюллетеня /29/. Модифицированная версия приводится в приложении 4.

Думается, опыт наших зарубежных коллег, уже решивших многие вопросы, которые только возникают перед отечественными разработчиками баз и коллекций машиночитаемых данных, будет чрезвычайно полезным для создающегося в России Архива исторических данных.

Приложение 1

Стандарт описания исследования (SSD),
принятый в архивах данных по социальным наукам.

SSD состоит из пяти разделов. Первый (информационный), помимо полного названия массива данных, содержит информацию обо всех исследователях и организациях, связанных с данным исследованием. Второй (содержательный) раздел кратко описывает цель и методику исследования (генеральная совокупность и единицы наблюдения, процедуры выборки и взвешивания данных, описания исходных и производных файлов), т.е. всю процедуру первичного исследования. Третий раздел (его можно назвать административно-техническим) чрезвычайно важен для целей вторичного анализа: это условия доступа потенциального пользователя к данным, информация о процедурах проверки корректности данных, выполненных как разработчиком, так и архивом, а также сведения о стандартных пакетах прикладных программ, которые можно непосредственно использовать при работе с массивом. Два последних раздела являются библиографическими: они дают информацию о соответствующих публикациях (отчетах, докладах), связанных как с первичным, так и со вторичным анализом данных, а также об основных стандартных социо-демографических параметрах респондентов, содержащихся в массиве данных.

Общая информация

- 001 - Статус исследования в архиве машиночитаемых данных
- 002 - Классификация темы исследования (например, 03 - история и демография)
- 003 - Ключевые слова
- 004 - Язык, на котором создано описание исследования
- 005 - Аннотация
- 007 - Название исследования

Идентификация массива данных

- 101 - библиографическая информация о документации
- 111 - архив, в котором хранятся данные
- 112 - архив, в котором данные собирались
- 121 - владелец
- 122 - дата поступления коллекции в архив
- 131 - автор (авторы) исследования
- 132 - создатель коллекции (может совпадать с автором)
- 141 - инициатор исследования
- 142 - финансирующая организация
- 199 - дополнительная информация

Аналитическая характеристика данных

- 201 - тема исследования (аннотация)
- 202 - тип данных
- 211 - единицы наблюдения (объекты)
- 212 - число объектов
- 213 - размерности массива данных
- 214 - полнота данных, хранящихся в архиве
- 220 - хронологический период
- 221 - характер динамических параметров
- 222 - описание параметров генеральной совокупности
- 223 - процедура выборки
- 225 - территориальные рамки исследования
- 231 - даты создания информационного массива
- 232 - метод сбора данных
- 233 - метод исследования
- 234 - способы минимизации потерь информации
- 235 - администратор коллекции данных
- 236 - характеристика условий сбора информации
- 241 - модификация данных (процедуры взвешивания информации)
- 299 - другие особенности массива данных

Возможности вторичного анализа

- 301 - представление данных
- 302 - стандартные пакеты, в которых возможна обработка
- 303 - возможные информационно-поисковые системы
- 304 - информация, хранящаяся в информационно-поисковой системе
- 305 - использованная классификационная схема
- 311 - язык (языки) документов
- 321 - операции по проверке данных, выполненные исследователем
- 322 - операции по проверке, выполненные архивом
- 331 - условия доступа к данным
- 332 - ответственный за доступ к данным
- 399 - другие условия повторного использования данных

Ссылки на соответствующие публикации/результаты/исследования

- 401-409 - публикации/отчеты авторов первичного исследования
- 411-419 - другие публикации (повторные исследования)
- 421-429 - неопубликованные статьи/отчеты
- 431 - результаты анализа (шкалы, индексы и т.п.)
- 441 - ссылки на связанные с данным исследования
- 499 - дополнительная информация

Имеющиеся в массиве стандартные социо-демографические параметры

- 501 - основные персональные характеристики (возраст, пол, семейное положение...)
- 502 - место рождения
- 503 - место жительства
- 504 - жилищные условия
- 511 - характеристики домохозяйства
- 512 - характеристики родительской семьи/домохозяйства
- 521 - место работы
- 522 - род занятий
- 531 - доход
- 541 - образование
- 546 - социальная принадлежность
- 551 - политическая ориентация
- 556 - религиозная принадлежность
- 561 - капитал
- 562 - наличие предметов длительного пользования
- 571 - "культурные" потребности
- 576 - участие в общественных организациях

599 - другие параметры

Приложение 2

Стандарт описания исследования,
принятый в Датском архиве исторических данных.

Разделы "Общая информация", "Идентификация данных", "Аналитическая характеристика данных", "Возможности вторичного анализа", "Ссылки на соответствующие публикации/результат /исследования" и "Имеющиеся в массиве стандартные социо-демографические параметры" полностью соответствуют стандарту SSD, приведенному в приложении 1. Кроме того, вводятся дополнительный раздел "Описание источника":

Исходная организация информации

- 601 - цель создания источника
- 602 - диапазон данных источника
- 603 - содержание источника

Текущая организация информации

- 611 - способы организации записей
- 612 - способы интерпретации записей. Изменения, произведенные при архивировании исходных данных
- 621 - реорганизация и трансформация данных источника
- 622 - частичное целевое сокращение информации источника
- 623 - возможная случайная потеря части информации источника
- 624 - последствия такой потери

Возможность доступа к источникам

- 631 - архив, в котором хранится оригинал источника
- 632 - ссылочная информация
- 633 - возможность получения копии источника
- 634 - библиографические ссылки на публикации источника
- 641 - библиографические ссылки на основные научные работы с использованием этого источника

999 - прочая дополнительная информация

Анкета, заполняемая разработчиком, состоит из двух частей. Часть А касается источника, а часть В - процедур извлечения из него машиночитаемых данных. Если данные извлекаются из нескольких

источников, части А и В заполняются для каждого из них, описываются соответствующие промежуточные файлы данных, обеспечивая независимое вторичное использование каждого. Кроме промежуточных, в архиве сохраняется и составной файл, послуживший основой исходного исследования. В этом случае для финального файла часть А, очевидно, заполняется только с целью простого перечисления привлеченных источников, тогда как часть В должна присутствовать полностью.

Часть А состоит из 4 разделов: обстоятельства создания источника, его структура, адаптация источника для его использования в исследовании и условия доступа к источнику.

А. Описание источникового материала
(заполняется для каждого источника)

- А.1. "Нормативная" организация источника
 - А.1.1. Цель создания источника
 - А.1.2. Круг объектов, описанных в источнике
 - А.1.3. Содержание источника
 - А.1.4. Хронологические рамки источника
- А.2. Реальная организация источника (заполняется в том случае, если источник создавался не в полном соответствии с нормативом)
 - А.2.1. Процедура создания записей источника
 - А.2.2. Исходная интерпретация записей источника
- А.3. "История" перевода источника в машиночитаемую форму
 - А.3.1. Реорганизация или трансформация записей
 - А.3.2. Намеренное частичное сокращение информации источника
 - А.3.3. Ненамеренная утрата части информации источника
 - А.3.3.1. Описание обстоятельств утраты информации
 - А.3.3.2. Результат утраты информации
- А.4. Условия доступа к источнику
 - А.4.1. Архивные атрибуты
 - А.4.1.1. Место хранения оригинала источника
 - А.4.1.2. Архивные атрибуты
 - А.4.1.3. Условия копирования источника
 - А.4.1.4. Библиографическая информация для опубликованных источников
 - А.4.2. Библиографические ссылки на основные научные работы, основанные на данном источнике

Часть В состоит из 5 разделов: время создания машиночитаемого массива: программное обеспечение, пригодное для работы с массивом;

физические параметры массива; его логические параметры и методы архивирования (кодирование, стандартизация и т.п.).

В. Трансформация источника в машиночитаемый вид

- В.1. Время создания массива
 - В.1.1. Начало
 - В.1.2. Конец (может быть предполагаемый)
- В.2. Использованное программное обеспечение
- В.3. Физические характеристики
 - В.3.1. Носитель информации
 - В.3.2. Тип компьютера
 - В.3.3. Операционная система
 - В.3.4. Размер и формат файла (файлов)
- В.4. Логические характеристики массива
 - В.4.1. Структура данных (плоский файл, база данных, свободный текст и т.д.)
 - В.4.2. Количество единиц наблюдения (для плоских файлов - число объектов и признаков, для баз данных - число файлов, записей, полей, для текста - число слов или строк)
- В.5. Используемые методы
 - В.5.1. Процедуры выборки
 - В.5.2. Кодирование и стандартизация
 - В.5.3. Связывание записей
 - В.5.3. Другое (указать)

Приложение 3

Стандарт описания коллекции исторических данных, принятый в
Голландском архиве исторических данных

- 1. Идентификация/Происхождение данных
 - 101 - название исследования
 - 111 - архив, в котором хранятся данные
 - 112 - архив, в котором данные собирались
 - 121 - владелец
 - 122 - дата поступления коллекции в архив
 - 131 - автор (авторы) исследования
 - 132 - создатель коллекции (может совпадать с автором)
 - 141 - инициатор исследования
 - 142 - финансирующая организация
 - 149 - цель исследования
 - 199 - дополнительная информация
 - 100 - библиографические ссылки

2. Схема исследования/Принципы анализа

- 201 - сфера исследования (археология, экон. история, демография...)
- 202 - ключевые слова (всего 2-6)
- 220 - хронологический период
- 221 - характер динамических параметров (временные ряды, тренды, панельное исследование)
- 202 - вид данных (переписи, статистика, тексты, ..)
- 203 - источники данных (опубликованные, неопубликованные...)
- 237 - происхождение источников
- 238 - характеристика степени сохранности источников
- 211 - единицы наблюдения (персоналии, хозяйства, территории ...)
- 212 - число объектов (в т.ч. пропуски, замены ...)
- 214 - полнота исследования (по отношению к генеральной совокупности)
- 222 - описание генеральной совокупности
- 223 - процедура выборки
- 224 - отклонения от выборочного метода (в случае нерепрезентативности)
- 231 - даты создания коллекции
- 232 - метод создания коллекции
- 233 - методика анализа
- 234 - способы минимизации потерь информации (реконструкция ...)
- 235 - администратор коллекции данных
- 236 - характеристика условий сбора информации
- 321 - операции проверки данных, выполненные исследователем
- 213 - размерности массива данных (число объектов, показателей...)
- 219 - исходный формат хранения данных (числовой, текст, SPSS, база...)
- 299 - дефекты массива данных

3. Условия вторичного анализа

- 301 - формат представления данных (см. 219)
- 302 - стандартные пакеты, в которых возможна обработка
- 306 - статус исследования (где хранятся данные)
- 311 - язык документации (описания, codebook, др...)
- 322 - операции по проверке данных, выполненные архивом
- 331 - условия доступа к данным
- 332 - ответственный за доступ к данным
- 399 - другие условия повторного использования данных

4. Ссылки на публикации/результаты/исследования

- 401 - публикации/отчеты авторов первичного исследования
- 411 - другие публикации (повторные исследования)
- 421 - неопубликованные статьи
- 431 - результаты анализа (банки данных, индексы...)
- 441 - ссылки на исследования, связанные с данным
- 499 - дополнительная информация

5. Справочные параметры

- 501 - наличие стандартных справочных данных (индексы цен, зарплаты, ...)
- 999 - вспомогательная информация

Приложение 4

Стандарт описания машиночитаемых данных
Архива машиночитаемых данных по российской истории

Описание массива данных служит основой для его оценки экспертной Комиссией Архива и публикации сведений о нем в Ежегодном Каталоге.

Документация должна включать три обязательных части: полное и краткое описание и приложения.

Полный вариант описания должен включать следующие разделы.

A.1. Авторское название массива данных

B.2. Информация о разработчиках (ф.и.о, место работы, домашний адрес, телефоны, Fax, E-mail)

C. Третий раздел представляет характеристику используемого исторического источника или источников. Сведения об источнике должны служить более ясному представлению об использованных первичных материалах. Раздел должен содержать:

C.3. Полное название источника

C.4. Указание реквизитов хранения, если источник архивный, или библиографической информации, если он опубликован

C.5. Определение хронологических рамок исторического периода, к которому относятся данные источника

C.6. Указание территории (вся страна, отдельные регионы, др.), которую характеризует данный источник

C.7. Указание автора (конкретно-историческое лицо, организации, участвовавшие в создании источника)

C.8. Целевое назначение источника при его создании

C.9. Качественная характеристика источника (типы таблиц, описание

ключевых показателей, др.)

С.10. Общая характеристика объема источника (число анкет, таблиц, страниц текста)

С.11. Характеристика пропусков в данных

Д. Четвертый раздел полного описания представляет характеристику данных, включенных в массив и программного обеспечения. В этом разделе необходимо указать, какое стандартное программное обеспечение использовалось. Если использовалось оригинальное программное обеспечение, указать какое. В раздел включаются:

Д.12. Описание структуры данных (формат данных: плоские файлы, база данных, электронная таблица, свободный текст)

Д.13. Оценка полноты перевода информации в машиночитаемый вид (все ли объекты и показатели источника использованы, если нет, то указать принципы отбора)

Д.14. Указание на то, использовалась ли кодировка показателей, если да, то указать системы кодов и приложить к документации полные листы кодировки

Д.15. Указание на возможность пополнения массива данных (другие источники, периоды, территории)

Д.16. Ключевые слова данных

Д.17. Объем информации в KB с указанием формата (ASCII, dBASE, LOTUS, др.)

Д.18. Тип ЭВМ, на которой создан массив

Д.19. Используемое программное обеспечение (например, СУБД и т.д.)

Д.20. Носитель информации (магнитные ленты, гибкие магнитные диски и т.п.)

Е. Последний раздел характеризует апробацию данных. В этом разделе указываются:

Е.21. Проблемы, для решения которых применялись эти данные

Е.22. Имеющиеся публикации по созданному массиву данных (публикации конкретно-исторического характера реферируются с указанием методов исследования)

Помимо полного описания разработчики должны представить краткое описание в объеме до 1 стр. текста для включения информации в каталог. В нем необходимо указать:

- название массива данных;
- информацию о его разработчиках;
- название источников;
- территориальные и хронологические рамки информации;
- описание структуры массива данных на уровне содержания файлов;

- объем информации в КВ с указанием форматов файлов (ASCII, dBASE, LOTUS, др.);
- указание программного обеспечения;
- тип ЭВМ, на которой выполнена работа;

Приложения: списки показателей и объектов, структура каждого файла, листы кодировки.

Вся документация (полное, краткое описания и приложения) вместе с файлами данных представляется в машиночитаемом виде (желательно на гибких дисках) в 2 экз. Кроме того, необходимо представить распечатки всей документации также в 2 экз.

ПРИМЕЧАНИЯ

1. Thorvaldsen G. The Preservation of Computer Readable Records in the Nordic Countries // History and Computing, Vol. 4. 1992, N. 3.

2. Гарскова И.М. Концепции банков информации в историкосоциальных науках (опыт США) // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории АН СССР, N. 3, октябрь 1991; Гарскова И.М. Современный опыт создания банков информации в историко-социальных науках // Россия и США на рубеже XIX - XX столетий (Математические методы в исторических исследованиях). М., Наука, 1992; Моисеенко Т.Л. Об использовании банков машиночитаемых данных по истории в новой зарубежной историографии // История СССР, 1985, N 5; Моисеенко Т.Л. Применение количественных методов и ЭВМ в зарубежной медиевистике. (Опыт создания машинных банков информации) // Математические методы и ЭВМ в исторических исследованиях. М., 1984.

3. Austin E.W. The ICPSR Historical Data Archive: What Can Be Learned from its First 25 Years? // Data, Computer and the Past. Proceedings of the Conference "Archiving and Disseminating Historical Machine Readable Data" / Doorn P., Kluts C., Leenarts E. (eds.). Cahier VGI N. 5. Vereniging voor Geschiedenis en Informatica Vitgenverij Verloren - Hilversum, 1992.

4. Tjalsma H.D. Describing "Other" Data Material // Historical Data Archives on Common Ground. Odense, DDA, 1993 - preprint.

5. Doorn P., Matthezing H. After the Flood: Archiving Electronic Records in the Netherlands // History and Computing, Vol. 4, 1992, N. 3; Zweig R. Virtual Records and Real History // History and Computing, Vol. 4, 1992, N. 3.

6. Thorvaldsen G. The Preservation of Computer Readable Records in the Nordic Countries // *History and Computing*, Vol. 4. 1992, N. 3.
7. Doorn P., Matthezing H. Op. cit.
8. Anderson S.J. The Future of the Present - The ESCR Data Archive as a Resource Centre for the Future // *History and Computing*, Vol. 4, 1992, N. 3. P. 195.
9. Anderson S.J. Op. cit.
10. Morris R.J. The Historian at Balshazzar's feast: a Data Archive for the Year 2001 // *Data, Computer and the Past. Proceedings of the Conference "Archiving and Disseminating Historical Machine Readable Data / Doorn P., Kluts C., Leenarts E. (eds.). Cahier VGI N. 5. Vereniging voor Geschiedenis en Informatica Vitgenverij Verloren - Hilversum, 1992.*
11. Morris R.J. Op. cit.
12. Thorvaldsen G. Op. cit.
13. Doorn P. Data is Sacred, Opinion is Free. The Netherland Historical Data Archive // *Data, Computer and the Past. Proceedings of the Conference "Archiving and Disseminating Historical Machine Readable Data / Doorn P., Kluts C., Leenarts E. (eds.). Cahier VGI N. 5. Vereniging voor Geschiedenis en Informatica Vitgenverij Verloren - Hilversum, 1992; Gonzalez P. Computerization Project for the "Archivo General de Indias" // Ibid.*
14. Anderson S. and Winstanley B. Review of Documentation Procedures at the ESCR Data Archive // *Historical Social Research*, Vol. 18, 1993, N. 1.
15. Anderson S.J. Op. cit.
16. Anderson S. and Winstanley B. Op. cit; Anderson S. and Winstanley B. Problems and Possibilities of a Historical Thesaurus // *Historical Data Archives on Common Ground. Odense, DDA, 1993 - preprint.*
17. Geda C.L. Data Preparation Manual. ICPSR, 1980.
18. Anderson S. and Winstanley B. Op. cit.
19. Morris R.J. Op. cit.
20. Marker H.J. Towards a Study Description for Historical Data Materials // *Odense, DDA-Nyt, 1986, N. 39; Marker H.J., Reinke H. and Shurer K. Making Sense out of Historical Documentation // Standartisation and Exchange of Machine- Readable Data in the Historical Disciplines / Haussmann F., Hartel R., Kropac I. and Becker P. (eds.). Craz, 1986; Marker H.J., Reinke H. and Shurer K. Information Requirements and Data Description in Historical Social Research: a Proposal // *Historical Social Research*, 42-43, 1987; Marker H.J., Reinke H. and Shurer K. Sources and data: description and documentation requirements in historical social research // *Standarization et echange des bases de donees historiques / Genet J.P.**

(ed.). Paris, 1988; Reinke H. Towards Standard for the Description of Machine-Readable Historical Data // Historical Social Research, 1981, N. 18; Reinke H. Datenbeschreibung und Datendokumentation in der Historischen Sozialforschung, Problembeschreibung und Empfehlungen für die Forschung // Datenbanken und Datenverwaltungssysteme als Werkzeuge historischer Forschung / M. Thaller (ed.). Historisch-Sozialwissenschaftliche Forschungen Band 20, St.Katharinen, 1986.

21. Marker H.J. Standard for Description of Historical Data Materials // Historical Data Archives on Common Ground. Odense, DDA, 1993 - preprint.

22. Van Hall N.S. Towards a Standard for Description of Historical Datasets // Historical Social Research, Vol. 14, 1989, N. 1; Van Horic R. Development of a New Standard: the Historical Dataset Description Scheme // Nederlands Historisch Data Archief 1: Eindverslag van een pilot-project / Doorn P.K., van Horik R., Touwen L.J. (eds.). Amsterdam, Swidoc/Steinmetzarchief, 1990.

23. Touwen L.J. The NHDA - DIY Program // Historical Data Archives on Common Ground, DDA, 1993 - preprint.

24. Touwen L.J. Op. cit.

25. Marker H.J. Standard for Description of Historical Data Materials // Historical Data Archives on Common Ground. Odense, DDA, 1993 - preprint.

26. Базы данных в СССР. Каталог. Ч. 1, 2. М., 1990.

27. Базы данных в России

28. Гарскова И.М., Осокина Е.А., Тяжелникова В.С. Информация Консорциума по базам данных в исторических исследованиях // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории АН СССР, N. 2, август 1991.

29. База данных "Аграрное развитие губерний Европейской России на рубеже XIX-XX вв." // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории АН СССР, N. 3, октябрь 1991; База данных "Пищевкусовая промышленность Европейской России в 1900 году" // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории Российской Академии наук, N. 4, декабрь 1991; База данных по "личным делам" бакинских нефтепромысловых рабочих "Товарищества нефтяного производства "Братья Нобель" (нач. XX в.) - "Рабочие" // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории Российской Академии наук, N. 5, март 1992; База данных по торговой и демографической статистике "Снабжение населения промышленными товарами и продовольствием в 30-е гг." // Информационный Бюллетень Комиссии по

применению математических методов и ЭВМ в исторических исследованиях при отделении истории Российской Академии наук, N. 5, март 1992; Поздеева И.В., Шалугина И., Бородкин Л.И. О разработке базы данных "Старопечатная книга" // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории Российской Академии наук, N. 4, декабрь 1991; Юмашева Ю.Ю. Краткое описание базы данных "Командармы" // Информационный Бюллетень Комиссии по применению математических методов и ЭВМ в исторических исследованиях при отделении истории АН СССР, N. 2, август 1991.

О СОВРЕМЕННОМ СОСТОЯНИИ И ПЕРСПЕКТИВАХ РАБОТЫ
С АРХИВОМ МАШИНОЧИТАЕМЫХ ДОКУМЕНТОВ

А.К.Соколов

Данный доклад представляет собой плод размышлений о проблемах, которые возникают в связи с накопленным опытом компьютеризации в отечественном источниковедении, вернее в той его части, что непосредственно соприкасается с интересами автора, много лет работающего над созданием баз данных по советской истории. Как представляется, компьютеризация в отечественной историографии сегодня находится как бы на распутье. В предшествующие годы основной упор в применении ЭВМ был проблемно- и методо-ориентированным, если пользоваться терминологией М.Таллера /1/. Более того, подобные установки настолько доминировали, что их влияние сказывалось на используемом категориальном аппарате и оценке проводимых работ /2/. Отсутствие подобных установок трактовалось большей частью негативно, что неизбежно вело к тому, что источниковедчески-ориентированные исследования автоматически оттеснялись на второй план, хотя за ними стояли подчас годы напряженного труда, достижений, ошибок и разочарований. Не случайно всякого рода новации в творческой лаборатории историка шли под флагом применения количественных (математических, математико-статистических) методов и ЭВМ. Главным считалось продемонстрировать эффективность какой-либо новой методики и значимость достигнутого с помощью ее результата. Этим определялся выбор признаков и показателей для анализа, а значит, и обращение с источником. Однако, по мере того как определялись возможности и ограничения отдельных методик, подобный подход стал понемногу исчерпывать себя. Выяснилось, что проблемы, которые предстояло решать исследователям, имели чаще всего сложный и комплексный характер, и их решение не сводилось к применению единственного метода, а скорее совокупности методов, каждый из которых основан на своей системе аргументации и системе фактов, извлекаемых из источников. Отсюда непреложно проистекал вопрос об универсальном характере последней как обеспечивающей множественность подходов и сохранение наиболее полного информационного богатства источников. Как выяснилось, современный уровень компьютеризации позволяет успешно продвигаться именно в этом направлении. Об этом говорит возникновение такой дисциплины, как историческая информатика, которую многие авторы рассматривают как своего рода компьютерное источниковедение /3/.

В современной исторической информатике большое место занимает решение вопросов об адекватной передаче текста источника с сохранением

не только его фактического, но и познавательного содержания, поиска путей структурирования источника, позволяющего отобразить его сложность и многозначность, и формализации тех исследовательских процедур, которые знаменуют специфику исторического исследования. На этих основаниях зиждется, например, создание компьютерной системы КЛИО, разрабатываемой немецкими учеными и предназначенной для извлечения сведений из источников, связывания отдельных записей, установления связей между структурированными данными и полнотекстовыми базами данных, картографирования, обработки изображений и т.д., причем безотносительно к конкретной исследовательской задаче. Значение подобной работы трудно переоценить и, видимо, с точки зрения развития исторической информатики этот путь является наиболее перспективным.

Вместе с тем, видимо, следует иметь в виду, что истории как науке присуща очень высокая степень специализации знаний, и чаще всего исследовательское поле ученого-историка ограничено определенными рамками и пределами своего видения проблем. Это означает, что армия историков, по крайней мере на ближайшее будущее будет разбита по подразделениям, и проблемная ориентация в процессе компьютеризации истории сохранится. В связи с этим встает ряд вопросов: в какой мере необходима универсализация исторической информатики для историков всех времен и народов; стоит ли "городить свой огород", если, например, потребности исследователя современной истории вполне удовлетворяются за счет развития новых информационных технологий вообще; не уведет ли историков в сторону от столбовой дороги информатизации выдвижение на первый план специфических особенностей исторических источников (неполноты, рассеянности, неупорядоченности, неструктурированности информации и т.п.) и построение особых экспертных систем при их анализе. При этом встает и главный вопрос, действительно ли историческое исследование, особенно по современной истории, имеет фундаментальное отличие от любого другого рода деятельности, подразумевающего обработку информации, как утверждает М.Таллер? /4/. Да, во многих науках, в том числе науках об обществе и человеке, базы данных создаются путем заполнения форм, т.е. выбором из совокупности данных только тех, которые подходят для проведения эксперимента, но историк всегда может предъявить претензии своим коллегам (конечно, имеются в виду лишь родственные науки) об отсутствии в их базах данных исторического контекста, надлежащей динамики, и в слишком вольном обращении с реальными фактами. Разумеется, при правильном подходе к этим вопросам можно избежать высказываемых опасений, и напротив, - существенно расширить и обогатить спектр происходящих в различных науках информационных процессов. В то же время очевидна и другая сторона дела - необходимость, как говорят, постоянно "быть в курсе" всех возможных направлений компьютеризации в историческом исследовании.

Без сомнения, в настоящее время в этой области обнаруживается гораздо больше общего, чем можно было ожидать от исследователей, которые шли разными путями в науке. Вместе с тем нельзя не остановиться на некоторых особенностях, вытекающих из опыта работы проблемной группы, созданной в свое время В.З.Дробижевым для обработки массовых источников по истории советского общества, а затем лаборатории по применению ЭВМ и количественных методов в архивоведении и исторических исследованиях. При этом используются отдельные наблюдения о работе наших коллег, занимающихся сходными сюжетами как у нас в России, так и за рубежом.

Ориентируясь на введение в научный оборот больших массивов исторических информации, независимо от того, на что делался упор: извлечение ли их из определенного источника сведений, хранящегося в архиве, или сведение их в одну систему из разных источников, исследователи очень быстро убедились в необходимости элементарной категоризации информации и максимального разнообразия при выборе систем ее кодирования - единственной на определенном этапе компьютеризации формы представления данных в электронном архиве. В результате происходило быстрое наращивание книг элементарных кодов (code-books) и их унификация. Сами эти книги по своему содержанию приближались к структуре электронной записи. Естественно, стал возникать вопрос о значительном упрощении этой процедуры и выработке способов автоматического кодирования информации. Дополнительным аргументом для элементарной категоризации сведений был сам процесс обработки обширных комплексов массовых источников, приложение коллективных усилий и разделение труда участников этой обработки. Требовалось отделить ввод информации, сделать его независимым от субъективных влияний отдельного человека, обеспечить надежность первичных данных и избежать ошибок ввода. Таким образом, в этом аспекте процесс пошел как бы в обратную сторону от создания универсальных экспертных систем или баз знаний. В конечном счете это выливалось в кодирование и шифровку всех сведений источника. Оставалось сделать последний шаг - устранить вообще операцию кодирования и перейти к буквальному воспроизведению источника в электронной памяти.

В принципе хорошо известно, что современные средства компьютерной технологии и разного рода дополнительные устройства позволяют осуществить эту операцию с максимальной продуктивностью, намного повышая человеческие возможности. Однако в нашем отечестве для историков это пока остается красивой мечтой, поэтому при осуществлении этой операции сегодня приходится исходить из конкретных реалий.

Оптимальный путь обработки тех массовых источников, которые по тем или иным причинам попали в поле нашего зрения, в настоящее время, как и прежде, - создание баз данных. Новшества, которые в данном случае

обеспечиваются, – это введение свободного текста в информационные поля и возможность постоянного изменения структуры этих полей. В принципе в начальной стадии работы над источником можно ориентироваться и на любой текстовый процессор, но нужно помнить, что это существенно усложнит этапы дальнейшей обработки и анализа источника. Отчетливо осознавая относительную ценность предлагаемого пути для развития исторической информатики, мы, тем не менее, считаем, что круг проблем, которые историкам новейшего времени предстоит решать в ближайшем будущем, более всего связан с использованием структурированной информации, содержащейся в характерных для этого периода массовых источниках, имеющих тенденцию к стандартизации и унификации. Важно, однако, чтобы структура полей диктовалась самим источником, а не исследователем. Необходимо добавить, что процесс обработки информации исторических источников по новейшей истории больше напоминает не "охоту за трюфелями", по определению Ле Руа Ладюри, а переработку огромных пластов "породы", чтобы по крупницам собирать нужные сведения, и уже из них строить системы знаний. Исходя из этих положений, производится выбор стандартного пакета создания базы данных, который должен удовлетворять нескольким условиям, а именно: быть максимально гибким и позволяющим вводить в компьютер достаточно большие объемы информации, т.е. не содержать существенно сдерживающих работу ограничений, и достаточно широко распространенным, дающим возможность легко переходить от одной системы представления и обработки данных к другой. Конечно, сам вид и характер источника может определить выбор соответствующего пакета: будут ли это по преимуществу текстовые записи, цифры, различные меры или же таблицы данных, столь распространенные в новейший период.

Конечно, после реализации указанных процедур мы мало чего достигаем, помимо воспроизведения содержания источника при минимальной потере информации. Вопрос о том, какие методы будут использованы при анализе данных, остается открытым.

Дальнейшая работа с созданными базами данных характеризуется активным вторжением историка. Если этого нет, то они вроде бы остаются на уровне "мертвой информации", подобной той, которая содержится в большинстве архивных фондов и которой еще не коснулась рука исследователя. Однако сходство это обманчивое. Ни одна сколько-нибудь значимая разработка массовых источников не предпринималась ради них самих, а служила некоей цели, явно или неявно выраженной в историографии. Нынешний ее этап, связанный со сменой парадигм в ученом мире, несколько принизил значение ранее проводимой работы, заставил несколько по-иному посмотреть на направления и перспективы создания баз данных. Но, как представляется, значение многих баз данных, равно как и источников, на основе которых они были созданы, сохраняется. И раньше базы данных, которые создавались нами, например – для анализа

социальных структур, плохо укладывались в жесткие каноны, которым исследователи обязаны были следовать, по крайней мере в двух аспектах: во-первых, они демонстрировали ущербность схематического марксистского классового анализа общества, показывали, что социальная дифференциация российского и советского общества была не столь глубокой и однозначной, как представлялось, далеко не сводилась только к отношениям собственности и месту в системе общественного производства; во-вторых, на основе анализа данных становилась очевидной многомерность социальных процессов и невозможность подходить к ним с какой-либо заранее заданной конструкцией. Особенно поразительными явились в данном отношении материалы регистрации представителей бывших эксплуататорских классов, которую советская власть провела осенью 1919 г. и часть сведений которой были преобразованы в архив МЧД. Выяснилось, что существующие в литературе представления об облике "имущих классов" России совсем не адекватны реальным экономическим и социальным отношениям /5/. Впрочем, о том же в сущности свидетельствовали результаты анализа данных в многочисленных исследованиях о рабочих промышленных предприятий, служащих госаппарата, работниках других общественных групп и, в особенности, в работах по социальной дифференциации российской деревни.

В настоящее время становится очевидным, что новые подходы к интерпретации получаемых в процессе обработки данных должны базироваться на более сложных и тонких методологических и концептуальных моделях, чем принято считать. В них, с одной стороны, должны найти отражение достижения современной общественной мысли в изучении социальной динамики общества, а с другой, - учтена специфика той исторической реальности, которая имела место в России и СССР. Только на таких условиях можно взять на вооружение, например, современные теории классового расслоения, средних слоев, групп и группового поведения, теорию элит и т.д. Только так, в частности, можно реализовать проект "Номенклатура в системе власти в СССР", "Комиссары и экономика" или какие-либо другие. Первые же наработки в этой области показывают, что в случае с "этой страной" надо постоянно иметь ввиду ее особенности и спецификацию разного рода применяемых новаций к ее историческим реалиям. Вряд ли, например, мы найдем в литературе адекватную концептуальную модель такому общественному организму, как "номенклатура".

Приводимое рассуждение означает, что в процессе работы над созданием баз данных по современной истории мы должны, по крайней мере на ближайшее будущее, исходить из следующих обстоятельств: 1) что проблемная ориентация в разработке массовых источников (и не только их) будет превалировать и этим будет определяться процесс создания баз данных и архивов МЧД; 2) что концептуальные построения (экспертные системы) будут возникать в процессе непосредственной обработки

источников и создаваемых баз данных; 3) что на первый план выйдут не столько вопросы технологии создания отдельных баз данных, сколько их объединения и конструирования интегральных информационных систем /6/.

С учетом данных обстоятельств в настоящее время строится работа той группы исследователей, к которой причисляет себя автор доклада. Эта работа как бы распадается на три относительно автономных этапа. Первый состоит в создании баз данных, по возможности адекватных самому источнику, не влекущих за собой невосполнимой потери первичной информации. Именно поэтому все последние базы данных, созданные или создаваемые нами (по составу делегатов XII, XVII, XVIII съездов РКП(б) - ВКП(б), по аппарату Наркомюста, по личным делам работников аппарата ЦК ВЛКСМ 1986-1991 гг., по формированию ВПК и военно-промышленной элиты в довоенный период, по семейным бюджетам рабочих и др.), содержат попытку полного воспроизведения содержания источника со всеми присущими ему атрибутами. В этой связи встает задача своего рода "раскодирования" тех информационных массивов, которые были созданы ранее и где кодирование выступало лишь средством шифровки элементарной информации (партийные переписи 1918 и 1922 гг., переписи служащих госаппарата 1918 г., анкетные данные делегатов съездов Советов, депутатов ЦИК СССР I-VII созывов, Верховного Совета СССР I, II, VI, XI созывов и др.). Все эти базы следует рассматривать как основу для решения источниковедческих задач и разного рода преобразований исходной информации.

Несколько особняком в этом ряду стоит вопрос создания базы данных по персональному составу СНК 1917-1940 гг. Строго говоря, ее нельзя назвать ни источниковедчески-, ни методо-ориентированной. Это тип мобилизационной базы данных, которые, по нашему мнению, будут играть в будущем все большую роль. Определенный аналог подобной базы данных - Soviet Data Bank, созданный в США историками Дж.Арчем Гетти и У.Чейзом на основе опубликованных в печати биографических сведений о советской элите 1917 - середины 1930-х годов. Однако в первом случае круг задач, особенно источниковедческих, оказывается куда более широким и сопряжен со множеством трудностей при их реализации. Мало того, что отсутствует какой-либо единый источник сведений о народных комиссарах, на использование сохранившихся документов личного свойства, в том числе официальных, накладываются строгие ограничения. Но даже при условии получения доступа к информации встает проблема привлечения дополнительных сведений из различных источников, определения их надежности, достоверности, сравнения, сопоставления и т.д. /7/.

Второй этап работы над созданными базами данных, как уже говорилось ранее, знаменуется активной ролью историка-исследователя, который строит на основе имеющихся сведений свою систему обработки информации. При этом он должен соединить в данной системе различные аспекты ее семиотического, семантического и прагматического содержания

да еще с учетом их исторического контекста. Решить сразу эти задачи невероятно трудно, поэтому, видимо, в настоящее время надо исходить из того, что каждый исследователь будет создавать свою экспертную систему обработки данных, разрабатывать свои методы регулирования и сжатия информации. Каждый создаваемый вариант определяется прежде всего исследовательской целевой установкой, но немаловажную роль при этом должен играть сам источник, который оказывает "давление" и на характер обработки материала, и на корректировку поставленных историком задач. При этом предполагается, что сам процесс методологических поисков будет способствовать выработке общих подходов и приемов исследования, которые реализуются на третьем этапе работы.

Он заключается в последовательном конструировании системы знаний, получаемых как на основе анализа и синтеза информации отдельной базы данных, так и комплекса связанных между собой информационных массивов. При этом мы исходим из наличия в источнике различных слоев информации, каждый из которых может быть "исчерпан" на том или ином этапе обработки. В результате таких последовательных преобразований, извлечения информации все более высокого структурного уровня, применения все более точных и строгих методов ее обработки соответственно расширяются наши познавательные возможности. Конечная цель - построение такой информационной модели, которая будет более или менее адекватна или изоморфна предмету исследования.

ПРИМЕЧАНИЯ

1. См.: Таллер М. Что такое "источнико-ориентированная обработка данных"; что такое "историческая информатика"? // История и компьютер. Новые информационные технологии в исторических исследованиях и образовании / Отв. ред. Л.И.Бородкин и В.Леворманн. Max-Planck-Institut fur Geschichte In Kommission bei Scripta Mercaturae Verlag. St. Katharinen, 1993, с.5.

2. См.: Гарскова И.М., Тяжельникова В.С. Создание комплекса реляционных баз данных путем полного перевода источника в машиночитаемый вид // Метод в историческом исследовании. Материалы Всесоюзной школы-семинара. Минск, 1991; Они же. О методике перевода статистического источника в базу данных (на материалах Всесоюзной партийной переписи 1927 г.) // Вестн. Моск. ун-та. Сер.8, история, 1992, N3, с.65-68.

3. Моисеенко Т., Свищев М. Изучение аграрной истории России последних десятилетий: перспективы "компьютерного источниковедения" // История и компьютер. Новые .., с.152.

4. Таллер М. Указ. соч., с.10-11.

5. Подробнее см.: Смирнова Т.М. Создание базы данных по материалам "регистрации представителей бывших эксплуататорских классов 1919 г." // Источниковедение XX столетия. М., РГГУ, 1993.

6. Интересную постановку этого вопроса дает Э.Костолански. См.: Костолански Э. Интегральная информационная система (представление исторической информации и знаний) // История и компьютер. Новые .., с.185-192.

7. Подробнее см.: Бонюшкина Л.Е. Проблемы создания банка данных по персональному составу СНК 1917-1940 гг. // Источниковедение XX столетия.

НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ ПРОГРАММА
"БАНК ДАННЫХ ПО РЕГИОНАЛЬНОЙ ИСТОРИИ: УРАЛ В XX веке"

Т.И.Славко

В настоящее время сотрудниками исторического факультета Уральского государственного университета совместно с исследователями других научных учреждений формируется банк данных по региональной истории: Урал в XX в.

С созданием специализированных проблемно-ориентированных банков данных связано комплектование и использование современной источниковой базы. В результате значительно расширяются границы изучения историко-социальных объектов, повышается доказательность и информационная отдача источников, становится возможной постановка тех проблем отечественной истории, которые до недавнего времени было невозможно раскрыть лишь на основе традиционных подходов.

Формирование банка данных сопряжено с исследовательскими поисками в различных направлениях. Прежде всего важна разработка многих теоретических и методических проблем социального измерения, изучения истории регионального развития, исходя из возможностей ее раскрытия на базе имеющегося или вновь вводимого источникового материала. Необходим детальный анализ изменений производственной инфраструктуры, оптимального территориального размещения промышленных комплексов и сельскохозяйственного районирования, сущности территориально-административного управления, а также выявление важнейших направлений созидательной деятельности человека, преобразований в социальном облике различных слоев населения на переломных этапах исторического развития.

Организация банка данных связана с выявлением и систематизацией в нем в первую очередь массовых источников, которые дополняются единичными фактами. Комплексность при их отборе заключается в создании такой информационной базы, которая позволила бы всесторонне исследовать поставленную проблему. В этой связи в рамках предлагаемого банка данных осуществляется серия исследований, посвященных источниковедческой характеристике массовых источников по тем проблемам отечественной истории, актуальность которых объясняется, во-первых, слабой изученностью, во-вторых, значительным вниманием и интересом с позиции отечественной и зарубежной историографии.

В настоящее время в процессе создания регионального банка данных получены определенные результаты. Создан ряд баз данных по отдельным конкретным историческим проблемам, которые являются составной частью всего банка. На их основе подготовлены либо находятся в стадии

написания документально-монографические исследования по ряду важнейших научных направлений.

Одной из актуальных является проблема изучения демографической ситуации в регионе, связанной с естественным движением населения (рождаемость и смертность), его половозрастной структурой, браками и разводами, миграцией. Интересные результаты в этой области могут быть получены на основе актов гражданского состояния - свидетельств о рождении и смерти, браках и разводах в условиях социалистического государства, а также метрических книг, существовавших до 1918 г. Несмотря на неоднократные изменения первичной формы этих документов, в итоге может быть составлена унифицированная анкета, содержащая сквозные признаки, встречающиеся во всех или в большинстве записей за длительные промежутки времени. Именно поэтому стало возможным комплектование на их основе баз данных.

К настоящему времени создана база данных "Демографические процессы в городах Среднего Урала в 60-80-е гг. XX в.". Она позволяет сделать вывод об общей направленности демографических событий, колебаний уровней рождаемости, смертности, брачности и разводимости, а также их структурных закономерностях и особенностях, характерных для населения городов Среднего Урала. Что касается дореволюционного Урала XX в., то пока разработана более локальная, хотя и не менее информативная, база на основе метрических книг Екатеринбурга начала века.

Сейчас формируется еще одна база для моделирования демографических ситуаций во всех городах Урала за период 20-50-х гг. В ее основу положены паспорта городов, которые позволяют описать тенденции естественного движения населения в зависимости от изменений индустриального облика города, в том числе в зависимости от численности предприятий и количества производимой продукции, ввода в действие новых промышленных и энергетических мощностей. Сведения, сосредоточенные в базе, позволяют раскрыть зависимость численности населения и тенденций развития сферы бытового обслуживания, городского и пригородного транспорта, жилищного фонда. Материалы базы данных в совокупности с другими базами дают возможность провести типологию городов, выявить отличительные особенности малых, средних и крупных городов, оценить влияние численности и естественного роста населения на количественный состав рабочих и служащих в различных сферах народного хозяйства. Ценность данной базы состоит также в том, что в ней содержится подробная информация об уровне обеспеченности населения яслями, детскими садами и школами, средними специальными и высшими учебными заведениями, культурно-просветительными учреждениями и др.

В круг важнейших проблем, связанных с анализом социально-экономического и политического развития страны 20-30-х гг., входит исследование социальной структуры различных слоев общества. Эти

проблемы непосредственно связаны с использованием целой системы репрессивных мер.

Составной частью насильственных методов явилось лишение избирательных прав. К лишенцам относились крестьяне из числа раскулаченных, лица, занимавшиеся торговлей в период нэпа, монахи и духовные служители религиозных культов всех вероисповеданий и толков, служащие и агенты бывшей полиции, а также особого корпуса жандармов и охранного отделения, все лица, прямо или косвенно руководившие деятельностью полиции, жандармерии и карательных органов, а также участвовавшие в контрреволюционной деятельности против советской власти, осужденные, в том числе репрессированные по политическим мотивам.

Рассекречивание новых комплексов документов, прежде всего личных дел лишенных избирательных прав, дает возможность подойти к исследованию данной проблемы на системном уровне. Личные дела настолько информативны, что позволяют сформировать ряд специализированных баз данных. Прежде всего, это процесс лишения избирательных прав как составная часть репрессивной политики государства с точки зрения механизма действия законодательных мер и попыток граждан оправдать себя перед обществом. Отдельно формируется база данных о лишенцах-раскулаченных с целью построения модели так называемого "кулацкого хозяйства", исследования изменений в социальной психологии крестьянства этого периода, а также изучение влияния раскулачивания на традиции крестьянского быта.

С помощью базы данных по лишенцам - бывшим торговцам, которая также формируется в банке данных, появляется возможность создания социального портрета частного торговца, а также раскрытия динамики развития и свертывания частной торговли в 20-е гг. Изучение социального портрета бывших белогвардейцев на основе созданной нами базы данных по материалам лишенцев из числа бывших белогвардейцев позволяет раскрыть демографическую структуру рядовых участников Белого движения, показать степень участия лишенцев в контрреволюционном движении и мотивы их перехода на сторону Советской власти. При анализе лишенцев из числа священнослужителей на основе соответствующей базы данных необходимо подчеркнуть важность сюжетов, связанных с их исторической ролью в формировании и поддержании нравственных устоев среди основного населения. При описании портрета лиц, прямо или косвенно участвовавших или руководивших деятельностью полиции, жандармерии и карательных органов (база данных только начинает формироваться), можно выяснить их образ жизни до и после революции, степень участия в гражданской войне, характер деятельности в условиях социалистического государства.

Базы данных по лишенцам дополнены традиционной подборкой документов, позволяющей восполнить недостающие сведения, глубже и

детальнее раскрыть те или иные проблемы. В связи с этим подготовлены рукописи документальных сборников: "Судьба раскулаченных спецпереселенцев на Урале в первой половине 30-х гг.", "Церковь на Урале в 20-е годы", "Исправительно-трудовые учреждения 20-х гг.", "Из истории Белого движения: Временное Областное правительство Урала".

В первом из них содержится комплекс проблемно-систематизированных источников, всесторонне иллюстрирующих суть "кулацкой" ссылки на Урале; во втором сосредоточены материалы, раскрывающие историю церкви на Урале в первое десятилетие советской власти, показывающие реальное воплощение процесса отделения церкви от государства, а также трагическую судьбу многих священнослужителей. Здесь приводятся документы по закрытию церквей и связанные с этим различного рода инциденты, показывается государственная политика в отношении церкви, а также служителей культа.

Третья работа посвящена истории функционирования исправительно-трудовых учреждений Уральской области в 20-е - начале 30-х гг. В ней описаны социальный облик заключенных, в том числе несовершеннолетних, условия их пребывания в местах заключения, подробно показаны сфера использования принудительного труда, расходы на содержание заключенных - питание, санитарно-гигиенические нужды, медицинское обслуживание и др. Специальное место отведено рассмотрению взаимоотношений между заключенными и обслуживающим персоналом. В свою очередь, эта работа существенно дополнена некоторыми сводными материалами имеющейся в региональном банке базы данных "Социальный облик заключенных Нижнетагильского концентрационного лагеря N 2 за 1921 г.", в которой содержатся сведения на каждого заключенного в отдельности по материалам регистрационных книг.

Сборник документов, посвященный Временному Областному правительству Урала, существовавшему с середины августа по 10 ноября 1918 г., раскрывает ряд проблем, относящихся к истории кадетско-меньшевистско-эсеровской партийной коалиции. Социальная программа правительства сильно отличалась от большевистской. Социалисты не признавали советской власти в форме диктатуры одной только партии коммунистов. Они допускали частную собственность и частное предпринимательство как определяющий фактор общественного прогресса. Под давлением кадетов - своих партнеров по правительственной коалиции - им пришлось поддержать лозунг продолжения империалистической войны. Наряду с этим из документов видно, что правительство отнюдь не собиралось ликвидировать все завоевания революции. Новая власть препятствовала денационализации крупных предприятий, твердо стояла за оставление у крестьян бывших частновладельческих земель, за сохранение советской системы социального страхования рабочих, за регулирование рыночных цен на продовольствие.

Перечисленные выше базы данных постоянно дополняются за счет материалов, комплектуемых в ходе изучения устной истории. Это воспоминания бывших спецпереселенцев, письма, документы и воспоминания участников Великой Отечественной войны и реэмигрантов из Китая, ныне проживающих на Урале.

Большой объем индивидуальных сведений о промышленно-производственном персонале содержится в карточках личного состава предприятий. Это один из важнейших источников, используя который можно проследить пути и этапы формирования, развития различных профессиональных групп трудящихся, динамику их состава, выявить взаимосвязь и взаимозависимость между разнообразными факторами, влияющими на их социальный облик. Что касается 1930-х - начала 50-х гг., то этот источник является одним из основных при исследовании вольнонаемного и принудительного труда (репрессированные, спецпереселенцы), а также роли иностранных специалистов в развитии отечественной индустрии. Данная проблема, являясь составной частью регионального банка, решается нами пока в одном направлении - создается база данных по трудовому использованию ссыльно-переселенцев на заводах, фабриках и рудниках.

Наряду с этим для более полного описания экономических и социальных последствий насильственных методов управления экономикой в настоящее время создается база данных по колхозам Урала середины 30-х гг., в основе которой сосредоточены материалы динамического обследования хозяйственного и финансового состояния колхозов за 1933-1934 гг., проведенного по обширной (более 1 тыс. признаков) программе Центрального управления народнохозяйственного учета.

Представляет большой интерес и является актуальным анализ образа жизни рабочих и крестьян в единстве их семейной, производственной, культурной и бытовой сторон, изучение условий их существования на различных этапах советской истории. Характеристика демографических, социально-экономических, культурно-бытовых процессов в системе через образ жизни рабочей и крестьянской семьи дает возможность получить более глубокие знания о разных слоях и социальных группах рабочих и крестьян, различающихся по своему социально-экономическому статусу. Для изучения этой проблемы необходимым условием также становится привлечение массовых источников, среди которых важнейшее место занимают первичные материалы бюджетных обследований семей служащих, рабочих и колхозников. Высокая информационная насыщенность этого источника, балансовая увязка всех показателей, систематизированность сведений и их сопоставимость на протяжении длительных этапов советской истории дает возможность осветить многие из перечисленных выше проблем.

Что же касается регионального банка данных, то уже создана база данных о социально-экономической структуре крестьянства первой

половины 60-х гг., проживающего на Среднем Урале. В ней держится информация о составе крестьянской семьи, описываются жилищные условия и структура хозяйства, денежные доходы и расходы, система потребления, бюджет рабочего времени. В стадии разработки находится база данных о влиянии Великой Отечественной войны на уровень жизни рабочей семьи Урала.

Социально-политические процессы, проходившие в стране в период деятельности Коммунистической партии, невозможно объективно рассмотреть без тщательного анализа социального облика коммунистов. Для изучения этой важнейшей проблемы в банке данных создается база "Социальный облик коммунистов 20-х гг. на Урале". Основу базы составляют первичные материалы Всероссийской партийной переписи 1922 г. и переписи вновь принятых членов ВКП(б) 1924 г. Заключенная в них информация весьма разнообразна и включает в себя демографические показатели, сведения о родителях, образовании, трудовой деятельности до и после революции, партийное и революционное прошлое, материалы об участии в стачках, забастовках, нелегальных акциях, военной службе, отношении к религии, связи с крестьянским хозяйством. Особое внимание уделяется переписи вновь принятых членов партии, отражающей изменения, происшедшие в результате массовых вербовок в партию в 1923-1924 гг., которые способствовали во многом упрочнению политической системы 20-30-х гг.

Важнейшей темой при изучении региональной истории является анализ основных закономерностей функционирования рыночных структур. В банке данных создана база "Торговля на Урале в начале XX в." на основе материалов Уральского торгово-промышленного адрес-календаря, справочного издания, выходившего с 1899-го по 1915 г. В него включено свыше 30 тыс. адресов торговцев и промышленников региона, содержится реклама торговых фирм и промышленных предприятий как Уральского региона, так и фирм других регионов, представленных на уральском рынке. Сведения базы позволяют восстановить ретроспективную историю маркетинга региона, исследовать инфраструктуру торговли, мелкого и среднего предпринимательства в начале века. Они дают возможность изучить ассортиментную номенклатуру производимой и реализуемой продукции, динамику цен на нее, рассмотреть формы организации предпринимательской деятельности (торговые дома, акционерные общества, кооперативы и др.), изучить налоговую политику центральных и местных органов управления в отношении мелкого и среднего предпринимательства. Материал базы раскрывает опыт работы органов местного самоуправления по формированию массового потребительского рынка, постепенной переориентации производственной базы региона на производство товаров массового потребительского спроса.

В банке данных нами предусмотрен также блок культурологических проблем. Формируется база по теме "Студенчество Урала 20-х гг.:

структура, образ жизни, система обучения", в основе которой лежат материалы обследования условий жизни студентов Уральского политехнического института, проведенного в 1926 г. В ней сосредоточены сведения, раскрывающие демографическую структуру студенчества, социальное происхождение студентов, бюджет времени, затрачиваемого на учебу и отдых, структуру питания, уровень доходов и расходов, сложности в системе обучения.

Кроме того, в рамках банка данных на основе паспортов школ и педагогических училищ создается база данных по народному образованию в первой половине 50-х гг., связанная с раскрытием его материально-технической базы, социальной структуры педагогических кадров и др.

Ведется работа по комплектованию базы данных "Массовое сознание сельского населения Урала в 20-30-е гг." на материалах частушек. Это один из самых кратких, но при этом очень емкий, по глубине отражения сведений, жанр. Его основное назначение не в описании каких-либо фактов или событий, а в выражении к ним определенного отношения. Обработка частушек с помощью новейших методик позволяет проникнуть в психологию людей. В частушках этого периода хорошо прослеживается уход с исторической сцены старых явлений и возникновение новых, отражаются конфликты между различными социальными слоями общества. В общественном сознании происходят значительные перемены, связанные прежде всего с изменениями в экономике, социально-классовых отношениях, в быту. Появляются мотивы, обусловленные вновь формирующейся системой ценностей, новыми взглядами, мировоззрением. Частушка как массовый источник позволяет оценить эти проблемы в тесной зависимости и взаимообусловленности.

Одной из сложнейших проблем истории 20-х гг., явившихся следствием империалистической и гражданской войн, голода и разрухи, оказалось беспризорничество. До сих пор отсутствуют специальные исторические исследования, посвященные данной теме. Наличие в областном архиве карточек беспризорников, составленных по состоянию на 1926 г., позволило сформировать базу данных "Беспризорничество на Урале в 20-е гг.". На каждого беспризорника имеются сведения, раскрывающие его социальный портрет и образ жизни до момента беспризорничества и в период его. Определенная группа показателей отражает пути выхода из состояния беспризорности.

Весьма важной при формировании банка данных является проблема разработки единых принципов структурирования сведений источников, что связано с обеспечением сопоставимости информации, содержащейся в них, ее однородности в плане исследования поставленных конкретно-исторических проблем. В результате формируется единая система признаков и их градаций.

Несмотря на сложность измерения социальных характеристик, разработаны определенные методики, позволяющие формализовать

разнообразную информацию и обеспечивающие тем самым комплектование банка данных за счет различных материалов.

Выделение признаков по массовым источникам, заданным в виде формуляров или в форме, близкой к ним, а также в сводках, практически не представляет значительных сложностей, поскольку эти признаки сформированы в процессе создания источника. Необходимо учитывать лишь изменения в наборе и содержании признаков, заключенных в разных конкретных источниках в зависимости от тех или иных временных промежутков или региона страны. В этом случае следует формировать унифицированную анкету, включающую общие для всех или для большинства анкет признаки.

Иначе выглядит решение проблемы измерения признаков при исследовании массовых объектов по нарративным источникам. Методика формирования массовых источников на основе нарративной информации предлагает сначала выделение признаков, сведения по которым сообщаются в источнике многократно. Сложность при этом заключается не только в количественном описании признаков, но и в определении их градаций в текстовой формулировке, что обусловлено спецификой каждого источника, особенностями словесных характеристик, системой их измерения. Затем сформированная анкета заполняется с учетом сведений каждой единицы, и полученная совокупность однородных анкет исследуется с помощью математических методов на уровне группировки, средних показателей, методов многомерной статистики, позволяющих определить структуру и типы социальных явлений или процессов, построить их модели.

Базы данных и методы математической обработки оформляются в виде пакета программ с соответствующей инструкцией для пользователя. Одна часть приемов связана с задачами использования банка данных в качестве архивного источника и предполагает поиск и сортировку конкретных материалов среди массы источников и выдачу их исследователю. Другая часть соответствует исследовательским функциям банка. Программа призвана обеспечить многоаспектный поиск, включающий разнообразные вопросы, который должен производиться на естественном языке и как можно более простыми для историка способами. Поэтому одно из главных требований, предъявляемых к каждой программе, - это ее простота и локальность. Алгоритм программы составляется таким образом, чтобы сохранялась возможность внести в него коррективы и расширить круга задач за счет новых методов сбора, обработки, хранения и передачи информации.

Публикация работ из серии "Банк данных по региональной истории: Урал в XX в." предлагает выпуск изданий нескольких типов. Прежде всего это традиционные документальные материалы по важнейшим проблемам программы, содержащие разнообразную информацию массового и единичного характера. Второй тип изданий связан с результатами машинной обработки массовых источников и будет представлен полученными на основе

историко-математического анализа сведениями в виде группировочных таблиц, информационных матриц взаимосвязей между признаками и т.д. Третью группу составят работы теоретического и методического характера, связанные с проблемами создания банка данных, четвертую - публикации справочного характера.

СОЗДАНИЕ БАНКА ДАННЫХ ПО ГЕНЕОЛОГИИ БЕЛАРУСИ: ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ

В.Л.Носевич

Опыт последних лет показывает, что перевод массовых источников в машиночитаемую форму постепенно становится магистральным направлением в деятельности историков-клиометристов в странах СНГ. В этой связи особенно важна взаимная информированность о проектах, реализуемых или готовящихся к реализации. Она позволит, с одной стороны, избежать дублирования при обработке сходных материалов, которые находятся в распоряжении новых суверенных государств. С другой стороны, крайне важно, чтобы базы данных, создаваемые разными группами исследователей, были сопоставимы между собой и основаны на аналогичных критериях.

Ниже предлагается информация о программе, подготовка к реализации которой сейчас начинается в Республике Беларусь по инициативе Комитета по архивам и делопроизводству при Совете Министров РБ. Основной ее целью является создание машиночитаемого архива семейной истории, в котором содержалась бы информация по генеалогии жителей Беларуси.

Постановка такой задачи вызвана несколькими причинами. С одной стороны, в республике сейчас остро ощущается необходимость возродить историческую память и этническое самосознание, которые сильно пострадали в предыдущие десятилетия. Для этого, в частности, необходимо восстановить диахронные связи, объединяющие ныне живущих с прошедшими поколениями. Компьютеризация архивных генеалогических источников значительно повысит их доступность для широких слоев населения.

С другой стороны, восстановление семейных родословий имеет большое медико-генетическое значение. Оно позволит проследивать наследственные заболевания, давать соответствующие рекомендации при вступлении в брак, как это делается в странах, где массовые генеалогические источники уже обработаны. В условиях последствий Чернобыльской катастрофы не менее важна и способность различать наследственные генетические дефекты от приобретенных в результате воздействия радиации.

К тому же компьютеризация является единственной возможностью преодолеть трудности, с которыми сталкиваются работники архивов при выполнении генеалогических запросов. При нынешней организации работы архивы способны удовлетворять порядка сотни таких запросов в год, что очевидно недостаточно для страны, имеющей десятиmillionное население и многоmillionную диаспору. Радикально изменить ситуацию способна только автоматизированная поисковая система. Она же позволит предотвратить

механическое разрушение документов в случае частого обращения к ним. Наконец, генеалогические источники являются бесценным материалом для исторических исследований, для проведения которых также необходима компьютеризация исходных данных.

Массив генеалогических источников, относящихся к территории Беларуси, можно условно разделить на две большие группы. Первую из них составляют материалы, относящиеся к представителям дворянского (шляхетского) сословия, вторую - относящиеся либо ко всем сословиям, либо только к непривилегированным.

В круг источников первой группы входят материалы, связанные со всеобщей проверкой дворянских прав, которая после присоединения Беларуси к Российской империи проводилась губернскими дворянскими собраниями и утверждалась Санкт-Петербургской Герольдией. Особую ценность в этой группе имеют дворянские родословные книги, в которых информация уже частично упорядочена.

Вторую группу составляют традиционные генеалогические источники, в первую очередь - метрические книги и материалы переписей населения (ревизские сказки). Они в равной мере относятся как к шляхте, так и к непривилегированным сословиям.

Метрические книги охватывают период с конца XVII века (лишь по единичным приходам) до начала XX века. Степень их сохранности сильно варьирует по регионам, лакуны разной величины имеются практически по всем приходам.

Ревизские сказки охватили территорию Беларуси по мере ее вхождения в состав Российской империи. Четвертая ревизия (1785 г.) проводилась на востоке Беларуси, пятая и последующие (1795, 1810, 1816, 1834, 1850, 1858 гг.) охватили всю ее территорию. Степень сохранности также различна. Наиболее полно сохранились материалы по Минской губернии, хуже всего - по Могилевской и Витебской.

Ревизские сказки особенно ценны тем, что приводят сведения о людях, выбывших со времени предыдущей ревизии (отданных в рекруты, бежавших, переведенных в другие населенные пункты), т.е. информацию, которая отсутствует в других видах источников. Вместе с тем они не всегда полны, точно указывают фамилии и особенно возраст. Поэтому для построения точной родословной желательно сопоставление данных ревизских сказок, метрических книг и других документов (при наличии таковых).

К числу таких дополнительных источников относятся инвентари помещичьих имений. В них содержатся только сведения о непривилегированных сословиях. Иногда эти сведения очень подробны (с указанием имен и возраста всех членов семьи), иногда называется только глава семьи. Особенно ценны последовательные серии инвентарей, которые имеются для некоторых крупных латифундий. В отдельных случаях

они охватывают вторую половину XVI века, т.е. период, практически не покрытый другими видами источников.

Общий объем охарактеризованных выше источников можно приблизительно оценить в десятки тысяч единиц хранения, миллионы единичных фактов. Объем информации должен измеряться, видимо, гигабайтами.

Для успешной обработки такого крупного информационного массива очень важен правильный выбор как стратегии, так и тактики. Технически возможны три варианта:

1. Создание именованного электронного каталога, облегчающего поиск необходимой записи. Непосредственное ознакомление с самой записью в этой случае должно осуществляться традиционными способами.

2. Создание каталога сопровождается сканированием всей информации в виде картинки. В этом случае найденные по каталогу записи можно было бы выводить для чтения на экран дисплея.

3. Полная формализация записей с последующим вводом всей информации в поля базы данных (через клавиатуру или, в случае унификации шрифта промежуточных карточек, через "читающий" сканер).

Очевидно, что первый вариант является самым экономичным и быстрым для реализации, но и наименее удобным для последующего использования. Второй вариант существенно повышает удобство пользователя при почти таких же трудозатратах. Его недостаток в том, что хранение информации в графической форме требует огромных объемов машинной памяти. Поэтому он реален только в случае хранения информации на лазерных компакт-дисках. Третий вариант оптимален с точки зрения пользователя, поскольку обеспечивал бы доступ сразу ко всей имеющейся информации, ее быстрый поиск и сортировку по любому признаку. В то же время его осуществление требует колоссальных трудозатрат. К тому же повышается вероятность ошибок, которые трудно выявить, поскольку потребитель не соприкасается с документом в его первоначальном виде.

Чтобы предотвратить ошибки, можно было бы использовать методику, которую применяет для аналогичных целей Генеалогическое общество Юты (США): каждая запись прочитывается независимо друг от друга двумя сотрудниками, а затем так же параллельно вводится в компьютер, с последующей сверкой обоих вариантов и внесением необходимых уточнений. В этом случае, однако, вдвое увеличиваются и без того огромные трудозатраты.

Независимо от избранной стратегии возможны и различные подходы к самому процессу обработки данных. Можно поочередно обрабатывать каждый документ в полном объеме, извлекая из него сразу всю информацию, которую решено извлечь. Возможна и "послойная" тактика, при которой обращение к документу происходит неоднократно. В первый раз считывается только самая общая информация - скажем, названия населенных пунктов, сведения о жителях которых имеются в источнике.

При втором обращении извлекаются сведения о каждой фамилии (с указанием листов, на которых она упоминается), и только затем – о каждом индивиде. Это позволяет сочетать ввод информации в компьютер с традиционными методами использования источников, причем уже введенные данные выполняют роль каталога по отношению к оставшимся.

В любом случае осуществление программы является вопросом многих лет, если не десятилетий. В начале пути трудно предвидеть те финансовые, кадровые, технические условия, которые могут сложиться в дальнейшем. Поэтому окончательный выбор какой-то одной стратегии и тактики на этом этапе вряд ли оправдан. Явно нерациональным представляется только крайний вариант: однократное обращение к каждому источнику с извлечением сразу всей содержащейся в нем информации, поскольку при таком подходе большая часть документов будет ждать своей очереди очень долго, оставаясь по сути выключенной из употребления. В остальном же необходимо начинать работу таким образом, чтобы она не исключала развития ни в одном из возможных направлений.

Исходя из этих соображений, Комитет по архивам и делопроизводству при Совете Министров РБ приступает к разработке комплексной программы "Радавод" ("Родословие"). На первом этапе основная работа разворачивается на базе Национального архива Беларуси, в котором хранится самая значительная часть генеалогической информации. Программа охватывает два главных направления: генеалогию шляхты и непривилегированных сословий.

В первом направлении ближайшей задачей является формализация по единой схеме сведений о каждом шляхетском роде в целом: фамилия, герб, время появления в источниках и имя родоначальника, локализация оседлости, наличие представителей, занимавших высокие государственные посты. Фиксируются сведения о каждом документе, в котором упоминаются представители данного рода. Эта работа будет вестись в тесном контакте с Российским родословным обществом, которое осуществляет сходную программу на базе фонда Герольдии в Российском государственном историческом архиве.

На этом же этапе предполагается готовить к публикации материалы отдельных, наиболее ценных источников (например, именной указатель шляхетских родов Беларуси, упомянутых в переписях войска Великого княжества Литовского в 1528 и 1567 гг. или гербовники родов отдельных уездов).

По мере накопления данных они будут вводиться. В данный момент Национальный архив ведет эксперименты с различными системами управления базами данных (СУБД), чтобы выбрать ту, которая оптимально отвечает целям данной работы, в том числе и с точки зрения возможностей дальнейшего расширения базы за счет сведений о персоналиях.

В отношении непривилегированных сословий первой задачей является создание общего географического указателя. Он должен отражать название населенного пункта согласно источнику (при наличии нескольких форм названия фиксируется все с адресовкой к одной, принятой за основную), его административную (губерния, уезд, для периода Речи Посполитой - воеводство, повет), конфессиональную (диоцезия, деканат, приход), хозяйственную (имение) принадлежность. Параллельно дается современное название и локализация (область, район, сельсовет). Далее приводятся ссылки на документы (фонд, опись, дело, лист), в которых имеются сведения генеалогического характера о данном населенном пункте. В дальнейшем предусматривается расширение базы данных за счет сведений о каждой фамилии, а возможно - о каждом индивиду, зафиксированных в этих документах. В этом случае будет вводиться информация о фамилии, имени индивида, его возрасте, месте жительства, обо всех родственниках, упомянутых одновременно с ним, а также иные сведения (причина смерти, мобилизация в рекруты и т.п.).

Возможно, уже на первом этапе будет проведена более детальная обработка нескольких типичных документов (метрических книг, ревизских сказок) - до указателя фамилий или даже персоналий. Это позволит отработать методику и выявить проблемы, связанные с расшифровкой записей и их вводом в компьютер, поскольку не всегда такие проблемы можно предвидеть заранее.

Есть надежда, что программа "Радавод" не останется узковедомственной, а со временем превратится в программу общенационального значения. Возможно, к ней подключатся Институт истории АН Беларуси, Министерство образования, общественные организации (в частности, Объединение белорусской шляхты). К обработке источников планируется привлечь силы добровольцев.

Необходимым условием успешного осуществления проекта представляется взаимодействие с архивными службами и научными учреждениями сопредельных государств, в первую очередь - Литвы и России, с которыми Беларусь долгие столетия имела общую государственность.

III. ОБРАБОТКА ИСТОРИЧЕСКИХ ИСТОЧНИКОВ

ЭКОНОМИЧЕСКИЙ БАРОМЕТР

Ю.П.Бокарев

1. История единых экономических индикаторов

История экономического барометра восходит к 20-м годам, когда в США, Англии и других странах стали рассчитываться обобщенные показатели, главной целью которых было предсказать изменение экономической конъюнктуры в ближайшем будущем. Рассчитывался экономический барометр и в России сотрудниками Конъюнктурного института Н.Д.Кондратьева.

Недостатком экономических барометров 20-х годов был способ определения единого показателя – обычно он рассчитывался как средняя геометрическая из множества частных индексов. Исключением был экономический барометр Гарвардского университета: для его расчета отбирались показатели, которые раньше других реагировали на изменение экономической обстановки. Такая работа ведется в США и сейчас.

Однако многие исследовательские центры исчисление единых экономических показателей прекратили. Во-первых, неясна содержательная сторона индексов, получаемых от простого агрегирования цен, денежной массы, оборотов торговли, промышленного производства и т.д. Что такой индекс означает? Какое экономическое явление, какой процесс он измеряет? На эти вопросы трудно ответить.

Во-вторых, исчисление единых экономических показателей было прекращено в связи с успехами макроэкономического моделирования. Макроэкономические модели описывают экономику как единую систему на основе определенной теории и статистического изучения динамики и взаимосвязей между экономическими показателями.

Одна из таких моделей была построена в 1950 г. Л.Клайном в соответствии с кейнсианской теорией по данным о развитии экономики США за период с 1921 по 1941 г. Модель состояла из 12 уравнений и 4 соотношений и предназначалась для изучения потребительского спроса и

совокупного дохода. Прогностические свойства модели Клайна оказались невысокими из-за ее упрощенности, а также теоретических и математико-статистических недостатков. Но в целом модель Клайна сыграла большую роль в деле изучения экономики, положив начало целому направлению макроэкономического моделирования.

Современные макроэкономические модели включают в себя сотни уравнений и соотношений, и для их создания требуются многие годы, а также ЭВМ с большим объемом памяти и быстродействием.

Недостатками макроэкономических моделей являются их жесткая детерминированность, привязка к определенной экономической теории и большие погрешности, получающиеся в результате ошибок аппроксимации и решения огромных систем уравнений.

В настоящее время эконометрическое моделирование заходит в тупик. Стараясь описать экономику возможно более детально, исследователи все больше дробят переменные. Например, одна функция спроса в модели Клайна со временем превратилась в десятки функций в Бруклинской модели, завершенной в 70-х годах. А ведь и десятки функций не могут описать всех склонностей и намерений потребителей. В результате, с одной стороны, создатели макроэкономических моделей испытывают нехватку статистических данных, а с другой стороны, приходится считаться с ограниченными вычислительными возможностями современных ЭВМ.

Поэтому все большее число исследователей начинают отходить от макроэкономического моделирования. Они возвращаются к старой идее агрегирования множества экономических переменных. Таким образом, история единых экономических показателей не кончилась. Исследователи вернулись к ним, но на новой методологической и методической основе.

2. Центральная идея

В начале 80-х годов, изучая экономический барометр Н.Д.Кондратьева с помощью современных математических методов, я заметил, что он лучше предсказывал бы изменение экономической ситуации, если бы рассчитывался с помощью методов факторного анализа.

Факторный анализ был создан американскими и английскими учеными Ч.Спирмэнном, Л.Терстоуном, С.Бартом, Р.Кеттелом и др. Очень быстро он завоевал признание немецких, французских, итальянских, швейцарских и испанских ученых. В нашей стране факторный анализ широко используется школой академика И.Д.Ковальченко.

В основе факторного анализа лежит та идея, что наблюдаемое в действительности множество взаимосвязей между экономическими показателями порождается небольшим числом латентных факторов, носящих не количественный, а качественный характер и потому неосредственно не измеряющихся. Например, связь между ценами, денежным обращением и

размерами товарооборота существует лишь в условиях свободного ценообразования. Введите твердые цены, и, как бы ни изменялась денежная масса, цены, естественно, изменяться не смогут, но появятся товарный дефицит и избыток денег у населения.

Свободны цены или нет - это зависит от определенной хозяйственной политики. Так в экономический барометр легко вводится хозяйственная политика - она определяет характер взаимосвязей между экономическими переменными. По экономическому барометру даже легче изучать изменения хозяйственной политики, чем по законодательным актам или интервью членов правительства. Как определить, на деле или только на словах проводится экономическая политика? Посмотрите на матрицу взаимосвязей и вы легко ответите на этот вопрос.

Однако можно двинуться еще дальше: так обработать матрицу взаимосвязей, чтобы получить количественные характеристики такого фактора, определяющего взаимосвязи между экономическими индексами, как хозяйственная политика. В результате мы получим новый индекс, в котором выражаются наиболее характерные черты данной народнохозяйственной системы. Величина этого индекса зависит, с одной стороны, от величины использованных при его расчете индексов, а с другой стороны, от того, какую роль играет каждый индекс (цены, денежная эмиссия, товарооборот, производство и т.д.) в экономической конъюнктуре.

Ясно, что при изменении народнохозяйственной политики изменяется и роль конкретных народнохозяйственных показателей. Например, в нашей стране до 1992 г. рост цен в целом стимулировал экономическую активность, размеры товарооборота и платежеспособный спрос. Это было связано с тем, что низкие цены увеличивали дефицит, делали неприбыльным производство, а избыток "горячих" денег на руках у населения позволял развиваться спросу при умеренном росте цен. После реформы Е.Т.Гайдара положение изменилось. Стремительный рост цен обесценил сбережения, прежнего избытка денег у населения уже не было, и потому дальнейший рост цен вел к падению спроса, а значит, и торговли, и производства.

Таким образом, в 1991 г. рост цен благоприятно влиял на экономическую конъюнктуру, а в 1992 г. - неблагоприятно. Поскольку все это учитывается при построении экономического барометра, значит, нам удалось соединить хозяйственную политику с экономической конъюнктурой, то есть сложить "тетку Дарью с белой козой".

Что же измеряет единый экономический показатель, построенный с учетом тесноты и направления взаимосвязей между экономическими переменными? Теперь на этот вопрос можно ответить: он измеряет степень благоприятности (или неблагоприятности) общей экономической обстановки.

3. Методологические трудности

Первоначально областью применения факторного анализа была психология. Ученые проводили тестирование, результатом которого были количественные данные об ошибках человека при воспроизведении им последовательности чисел и фигур, о скорости ответов на вопросы экспериментатора, о степени согласованности их с ответами других участников эксперимента и т.д. Измерив взаимосвязи между этими количественными данными и обработав полученные результаты, ученые получали возможность количественно измерить такие сугубо качественные характеристики, как память, осторожность, подчиненность коллективу и т.д. Субъективно мы понимаем, что у одного человека память лучше, а у другого – хуже. А насколько хуже? Ответ на этот вопрос дает факторный анализ.

Постепенно факторный анализ стал использоваться и в других науках, в том числе и в экономике. И вот здесь сразу возникли определенные трудности. Одно дело характер человека и совсем другое дело особенности экономики.

Черты нашего характера довольно консервативны. Поэтому матрица взаимосвязей, построенная путем обработки психологических тестов, достаточно устойчивая. Экономика же постоянно изменяется. Поэтому устойчивые взаимосвязи между экономическими переменными встречаются довольно редко. Значит, факторная обработка корреляционной матрицы будет отражать только текущее состояние экономики, но не ее прошлое и будущее.

К счастью, направления изменения тесноты взаимосвязей между переменными можно предсказать. Для этого надо рассчитать коэффициенты корреляции методом скользящих средних. Вы увидите, что в одних случаях будет наблюдаться тенденция к усилению или ослаблению взаимосвязей во времени, а с других случаях теснота взаимосвязей будет циклически изменяться.

Вторая сложность заключается в существовании лагов. То есть состояние одной переменной влияет на состояние другой переменной не сразу, а спустя какое-то время. Кроме того, эта вторая переменная тоже в свою очередь может оказывать обратное воздействие на первую, но через другой промежуток времени. Не остается никакого другого выхода как рассматривать одну и ту же переменную в разные моменты времени в качестве различных переменных, хотя это в значительной степени затрудняет расчеты.

Третья сложность заключается в том, что расчету коэффициентов корреляции между временными рядами препятствует тренд. К сожалению, мы не можем избавиться от него традиционным способом – с помощью параболических кривых или скользящих средних, так как при этом мы выбрасываем за пределы анализа характер долговременной экономической

динамики. Поэтому я был вынужден решить проблему тренда иначе, путем перехода к новым переменным на основе исчисления конечных разностей. Например, вместо динамики цен я рассматриваю скорости изменения цен и т.д. Такой подход имеет то преимущество, что от скоростей изменения экономических переменных можно легко перейти к динамике переменных, складывая соседние уровни. Таким образом, экономический барометр учитывает и долговременные тенденции.

4. Программа "Экономический барометр"

Для построения экономического барометра мной используется 120 переменных, получаемых по корреспондентским данным. Эти переменные непосредственно заносятся в программу, написанную мной в 1989 г. на языке GW-Basic. Программа производит следующую работу.

1. Переводит переменные из натуральной формы в индексы с базой на январь 1989 г. Агрегирует исходные переменные в 12 обобщенных индексов: цены, денежное обращение, курс рубля, кредит, фондовый рынок, внутренняя торговля, промышленное производство, строительство, транспорт, занятость, налоги.

2. Рассчитывает конечные разности.

3. Определяет взаимосвязи с учетом лаговых соотношений до трех месяцев, выбрасывает из матрицы закоррелированные переменные (при сохранении в ней представителей всех 12 исходных индексов).

4. Обращает переменные с отрицательными связями. При этом принимается во внимание место в структуре взаимосвязей некоторых переменных. Например, обращение внутренней торговли не допускается, иначе барометр может изменить направление оценки экономической обстановки.

5. Вычисляет первый главный фактор и рассчитывает остаточную корреляцию.

6. Вычисляет скорости изменения экономического барометра и сам экономический барометр.

7. Прогнозирует состояние каждого из 12 индексов и экономического барометра в целом на три месяца вперед.

5. Апробация

Первые результаты расчетов были опубликованы в ноябре 1991 г. в газете "Деловой мир". С декабря 1991 г. по сентябрь 1993 г. барометр систематически печатался в газете "Биржевые ведомости". В настоящее время срочная публикация экономического барометра прекращена из-за разрушения корреспондентской сети, но расчеты продолжают вестись по данным официальной статистики, поступающим с 2-3 месячным опозданием.

Экономический барометр позволил в ноябре 1991 г. предсказать наступление длительной депрессии фондового рынка в России, начавшейся с декабря; в декабре 1991 г. - резкое ухудшение общей экономической обстановки с января 1992 г.; в январе 1992 г. - финансовое ухудшение, начавшееся с февраля и приведшее к триллионной задолженности предприятий и т.д. Из месяца в месяц он давал верные месячные прогнозы как общего экономического развития страны, так и отдельных сторон народного хозяйства. Например, прогнозы инфляции давались им с точностью до 3%; курса рубля - до 5%; промышленного производства - до 4%.

К сожалению, более длительные прогнозы (например, трехмесячные) оказались менее точными, хотя в целом правильными. Я полагаю, что главная причина этого - резкие колебания в проведении экономической политики, влекущие за собой значительную перестройку системы взаимосвязей между переменными. Например, в марте 1992 г. правительство значительно смягчило денежную политику, и взаимосвязь между денежной массой, курсом рубля, торговлей и производством ослабла.

В настоящее время ведется работа по учету последствий для прогноза резких изменений тесноты взаимосвязей между переменными. Если эту работу удастся выполнить, то можно будет использовать экономический барометр и для долговременных экономических прогнозов (в форме вероятных сценариев).

СИСТЕМА ДЛЯ РАБОТЫ НАД МАТЕРИАЛАМИ ПОДВОРНЫХ ОПИСЕЙ

В.Л.Акимов, Н.М.Арсентьев

Прогнозирование массовых явлений и процессов – одна из центральных задач общественных наук XX века. Разработка этого направления отражает реальные запросы и практические потребности современного быстро изменяющегося общества. В такой ситуации вполне естественным кажется стремление историков осмыслить не только современный, но и прошлый опыт динамики исторических процессов на протяжении длительного периода времени. Осуществление таких исследований невозможно без тщательного научного анализа развития народонаселения. Уникальный материал в этом плане дает история развития социальной структуры российского общества XVIII – первой половины XIX века. Основным и хорошо сохранившимся источником демографической информации для этого времени являются материалы ревизий, то есть учета податного населения. Ревизии (от латинского *revisio* – пересмотр) проводились периодически. Всего было проведено 10 ревизий (1-ая – 1719, 2-я – 1744 – 1745, 3-я – 1763, 4-я – 1782, 5-я – 1795, 6-я – 1811, 7-я – 1815, 8-я – 1833, 9-я – 1850, 10-я – 1857). Ревизия продолжалась 1-2 года, а затем до следующей шла так называемая длительная ревизия, в которой уточнялись полученные сведения. Единицей учета мужского населения была "ревизская душа", данные о которой вносились в именные списки – ревизские сказки. Материалы ревизий дают общую численность населения России по полу, возрасту, семейному состоянию, занятиям, содержат также сведения о сословном и национальном составе. Иногда ревизии назывались "народными переписями", потому что не охваченная ими доля населения не превышала 5% всех жителей страны. Этот универсальный источник всегда привлекал к себе внимание исследователей /1/. Возможности, которые предоставляет персональный компьютер для обработки ревизских сказок, позволяют ускорить введение этих материалов в научный оборот. Дальнейшее развитие региональной демографии на современном этапе просто невозможно без компьютеризации исторических исследований. В этом плане прежде всего необходимо создать новые, более современные системы хранения, обработки, анализа материалов ревизий и сделать их доступными для многократного использования. Опыт работы группы исследователей из Мордовского университета может помочь в реализации этих задач.

1. С чего мы начинали?

Наша исследовательская группа возникла в середине 80-х годов. Она объединила ученых разных специальностей. Предметом наших интересов

стали процессы становления и развития социальной структуры промышленного общества в конце XVII – первой половине XIX века. Изучение этого вопроса проводилось на материалах Замосковского горного округа, одного из старейших центров металлургической промышленности страны (Владимирская, Калужская, Рязанская, Пензенская, Нижегородская, Тамбовская, Орловская, Тульская губернии). Итогом нашей работы стало учебное пособие "ЭВМ в историческом исследовании", изданное в 1989 году /2/. В нем была приведена программа для ЕС ЭВМ по обработке подворных описей. Анализируемые нами подворные описи были значительно сложнее приведенных в ревизских сказках, потому что они содержали данные о хозяйственном состоянии двора. Однако в отношении имеющих в них демографических сведений все типы подворных описей дают в основном однородный материал. Из документов, в которых зафиксированы подворные описи заводов Замосковского горного округа, наряду с материалами ревизий, следует выделить дела, объединенные под названием "Об истребовании от заводчиков заключения по составленному положению для Белорицкого завода, по применению оною к их заводам". Они появились после предписания Департамента Горных и Соляных дел от 3 декабря 1817 года. Тогда и сформировался целый комплекс однотипных источников /3/. Среди других архивных документов, содержащих подворные описи, следует еще выделить заводские ревизии. Они проводились в связи с продажей предприятий, а также разделом имений между наследниками. Именно так возникли проанализированные нами подворные описи 1838 года по заводам и селам наследников А.Р.Баташева /4/.

Запись двора, к примеру Виндреевского завода за 1818 год имела следующий вид: " 1/ Иван Яковлев – 57 лет. У него дети: Федор – 37 лет, Григорий – 25 лет, Антон – 23 года. У Федора жена Василиса – 36 лет. У них дети: Яков – 12 лет, Прасковья – 14, Аксинья – 8. У Григория жена Степанида – 27 лет. У них дочь Татьяна – полугоду. У Антона жена Агафья – 21 года. Итого мужского пола – 5, женского пола – 6. У Ивана Яковлева дом: изба квадратной меры 9 аршин, бревны в срубе 6-ти вершков, в ней окон – 3, при ней сени бревенчатые, в них два чулана, забранные досками. Против избы горница одинаковой с избой меры, в ней окошек 2. В них полы и потолки пластичные, при сенях крыльцо со входом. Все оное соснового лесу, крыты по лубку дранью. На улицу ворота и калитка с дубовыми столбами, крыты тесом. Возле ворот по линии амбар длиною 7 аршин, с навесом одинаковой меры. При нем пристенок и вообще с амбаром крыты по лубку дранью. Погреб с погребницею в квадрате по 6 аршин. Крыты таким же образом. Все оное строение соснового лесу. Двор длиною 15 сажень, шириною – 8 сажень. Кругом всего двора, кроме означенного строения, сделаны на столбах сараи с подъездом и крыты одной дранью. Позади двора огород длиною 21 сажень, шириною 8 сажень. Огорожены разным частоколом. Позади огорода баня с предбанником соснового леса, длиною 10 аршин, шириною 6 аршин,

крыты дранью. Иван Яковлев имеет домашнего скота: 1 лошадь, 3 коровы, 2 подтелка, 10 овец, 6 свиной с поросятами, 11 кур. Сена накашивают в год до 720 пудов" /5/.

Данные о дворах, содержащиеся в материалах по другим заводам, имеют сходную форму. Опыт эксплуатации программы обработки подворных описей на ЕС ЭВМ показал прекрасные возможности. Однако были выявлены и весьма существенные недостатки, которые в то время так и не удалось преодолеть, так как они были связаны с техническими средствами существовавших вычислительных машин.

Основным препятствием на пути широкого использования ЭВМ для обработки подворных описей стала сложность ввода исходных данных, которые почти вслепую кодировались на перфокартах. Это приводило к появлению большого количества ошибок и необходимости многократного последующего контроля введенной информации, что для массового источника весьма трудоемкое занятие.

Появившиеся вскоре диалоговые средства операционной системы ОС ЕС несколько упростили процедуру ввода и корректировки данных, однако возникли новые, не менее неприятные проблемы, связанные со сложностью работы в самой среде операционной системы. Дело в том, что операционная система ОС ЕС была рассчитана в основном на специалистов, имеющих достаточно высокую подготовку в области программирования и вычислительной техники и специалист с гуманитарным образованием должен был проходить довольно долгий и малопонятный ему курс обучения, что обычно сразу же отбивало у него желание работать на ЭВМ.

Указанные обстоятельства породили устойчивое мнение, что применение ЭВМ для обработки массовых источников занятие весьма полезное и перспективное, но чрезвычайно сложное и потому не получившее действительно широкого распространения. Появление и массовое распространение персональных компьютеров позволило перейти на новый качественный уровень в решении этой задачи. Технические средства и программное обеспечение современных персональных компьютеров ориентированы в первую очередь на массового пользователя, имеющего минимальный объем знаний о вычислительной технике и программировании. Современная идеология программирования для персональных компьютеров предполагает создание программных средств для работы с которыми не требуется длительной подготовки и обучения.

В 1993 году авторами была разработана новая версия программы для обработки подворных описей, названная РО. Программа предназначена для использования на персональных компьютерах, совместимых с IBM PC. Реализация самих методов группировки не представляет собой сложности, поэтому основной упор при разработке программы был сделан на создание максимальной простоты и удобства работы с программой для пользователя.

2. Состав и назначение программы РО

Программа обработки подворных описей РО представляет собой интегрированную оболочку, включающую средства для ввода исходных данных, их обработки по нескольким алгоритмам, просмотра и корректировки результатов. Для удобства пользователя управление интегрированной оболочкой осуществляется через разветвленную систему меню с обязательными подсказками в нижней части экрана.

Ввод и вывод информации осуществляется через систему файлов. Результаты подворных описей заносятся в файлы исходных данных, которые имеют стандартную форму. В первой строке файла располагается наименование объекта подворной описи /название деревни, завода и т.д./, затем следуют записи описаний отдельных дворов. Принятая стандартизированная форма записи файлов исходных данных позволяет легко производить практически любые преобразования исходной информации, например слияние файлов для совместной обработки, их экспорт в системы управления базами данных, создание банка данных и т.д.

Вывод результатов группировки осуществляется в обычные текстовые файлы в виде готовых печатных форм, которые в дальнейшем могут быть выведены на печатающее устройство с помощью любого текстового редактора, с которым привык работать конкретный пользователь. Отказ от ввода результатов непосредственно на принтер объясняется большим разнообразием используемых в настоящее время печатающих устройств, поэтому включение в программу набора драйверов для них привело бы к существенному ее усложнению. Кроме того наличие файла результатов с печатной формой создает для пользователя ряд дополнительных удобств: возможность создания архивов результатов и произвольного видоизменения таблиц перед выводом на печать, получение неограниченного количества копий и т.д.

Для ввода и корректировки исходных данных в программе РО имеется специализированный редактор. Информация вводится и редактируется с помощью специального экранного бланка, вид которого показан на рис. 1. При вводе количественных показателей контролируются диапазон вводимого числа и правильность представления, например, попытка ввода слишком большого числа или буквы "0" вместо нуля блокируется и программа предлагает повторить ввод снова. Качественные показатели "Тип семьи" и "Род занятий" вводятся через специальные меню, а затем автоматически кодируются. Все это позволяет существенно снизить вероятность ошибок при вводе данных и упрощает работу пользователя. При необходимости введенная информация может быть откорректирована.

Подсистема обработки имеет набор средств для выбора и проверки файла исходных данных и формирования файлов результатов. Файл исходных данных выбирается с помощью специального меню, напоминающего Norton-

Commander. Затем производится проверка выбранного файла на наличие ошибок. Проверяется структура записей и непротиворечивость количественных показателей. При обнаружении ошибок автоматически формируется файл ошибок, а на экран выводятся предупреждающее сообщение. Имя файла результатов формируется автоматически из имени файла данных, а если такое имя уже существует пользователю предлагается задать новое имя самостоятельно.

Программа обработки подворных описей реализует набор стандартных видов группировки и группировка по произвольной совокупности показателей /по наличию скота или земли во дворе и т.д./. К стандартным относятся следующие виды группировки: - половозрастной состав и половозрастная пирамида; - поколенный состав; -типология семьи.

В результате этих группировок формируются печатные формы стандартного вида, показанные в приложении /см. прилож. 1-3; рис.2/.

Наибольший интерес представляет группировка по произвольной совокупности показателей, которая позволяет получать таблицы результатов любого необходимого вида. Ограничения накладываются только на ширину таблицы /не более 20 столбцов/, так как нельзя обрабатывать и печатать строки длиннее 255 символов. Но и это ограничение может быть снято, если разбить широкую таблицу на ряд более компактных.

В программе реализованы два метода задания вида таблиц: с использованием готовых табличных форм и формирование произвольной таблицы в режиме диалога. Готовые табличные формы хранятся в специальных файлах, определяющих внешний вид таблицы и алгоритм группировки. Создание новых форм осуществляется с помощью встроенного генератора табличных форм. Процесс создания новой формы полностью автоматизирован и занимает несколько минут. В том случае, если пользователь не хочет создавать новую табличную форму, а имеющиеся формы его не устраивают, он может задать вид таблицы и алгоритм группировки самостоятельно в диалоговом режиме. Такой режим работы используется в случае создания одиночных или промежуточных таблиц, например, при проверке каких-либо гипотез, когда окончательный вид таблицы еще не ясен.

В качестве примера произвольной группировки в приложении приведена таблица дифференциации семей по количеству скота во дворе /см. прилож. 4/.

Для группировки по столбцам могут быть использованы произвольные простые или составные признаки. Под простым признаком понимается любой одиночный показатель описи, группировка по которому производится при падении его значения в заданный интервал или при равенстве заданному значению /например, столбцы 3-7 прилож.4/. Составной признак представляет собой совокупность нескольких простых признаков объединенных между собой с помощью логических функций AND /логическое

И/ или OR /логическое ИЛИ/. В качестве примера составного показателя рассмотрим второй столбец прилож. 4. Группировочный признак "Наличие мелкого скота во дворе" формируется из простых признаков "Количество овец" и "Количество свиней", объединенных логической функцией "ИЛИ". Группировка по этому признаку производится в том случае, если значение хотя бы одного из простых признаков больше нуля.

Группировочные признаки располагаются по столбцам в порядке возрастания их важности для дальнейшего анализа, так как один и тот же двор может одновременно попадать под действие разных признаков. Так например, во дворе одновременно может быть мелкий скот, коровы и лошади, но количество лошадей в данном случае более важно для анализа состояния двора, поэтому группировка будет происходить по этому признаку.

Последние три столбца таблицы результатов формируются автоматически для любых таблиц. В первом из них помещается информация о несгруппированных данных, то есть о тех дворах, которые не попадают под действие ни одного из группировочных признаков. В двух других столбцах приводятся итоговые и усредненные результаты.

Строки таблиц результатов прилож. 4 отражают все введенные в файл данных количественные показатели подворной описи, то есть таблица отражает все возможные взаимосвязи группировочных признаков с количественными показателями подворной описи. Возможно в некоторых случаях эта таблица содержит избыточную информацию. Однако этот недостаток нельзя признать существенным, так как удаление лишних строк с помощью текстового редактора занимает очень мало времени. В программе имеется встроенный текстовый редактор совместимый с широко известным редактором WordStar. С его помощью пользователь может, не выходя из интегрированной оболочки, просмотреть таблицы результатов группировки и при необходимости откорректировать их.

Варианты использования количественных показателей прилож. 4 могут быть самыми разнообразными и не ограничиваться только методом группировок. Можно решать задачи по изучению взаимосвязи явлений. К примеру, была подтверждена зависимость численности семьи от наличия земли.

В дальнейшей работе по обработке подворных описей объем вводимых в базу данных количественных показателей будет увеличен. Мы попытаемся максимально использовать всю информацию, представляемую данными источником. Уже находится в стадии изучения проблема сложения комплексов жилища промышленных рабочих. Без характеристики жилища и хозяйственных построек, их состава нельзя полно представить образ жизни заводского населения.

Другими словами, пока еще мы находимся в самом начале реализации тех многогранных возможностей, которые представляет нам персональный компьютер в историческом исследовании.

ПРИМЕЧАНИЯ

1. См.: Рашин А.Г. Население России за 100 лет /1811-1913 гг./. Статистические очерки. М., 1956; Кабузан В.М. Народонаселение России в XVIII-первой половине XIX в. /По материалам ревизий/. М., 1963 и др.
2. Арсентьев Н.М., Малкин В.И., Якупов З.Я. ЭВМ в историческом исследовании. Саранск, 1989.
3. ЦГИАМ, ф.2199, оп.1, д.108-114а.
4. Там же, д.273-279.
5. Там же, д.110, л.37-38.

АГРАРНАЯ ПОЛИТИКА ГОСУДАРСТВА
В ФОКУСЕ ОБЩЕСТВЕННОГО МНЕНИЯ СЕЛЯН В НАЧАЛЕ 1920-х ГОДОВ
(опыт работы с программным обеспечением SPSS/PC+)

Д.Х.Ибрагимова

Уникальность ситуации периода нэпа заключалась в смене государственно-централистской модели экономических ориентиров общества, вызвавшей известное раскрепощение социальной и духовной сфер. В аграрном секторе можно выделить две основные тенденции, порождавшие сложные метаморфозы в сознании сельского населения: движение к "американизации" крестьянских хозяйств и развитие кооперации. В свою очередь, взгляды, установки, настроения селян, будучи специфическим отражением действительности, явились и побудительными мотивами их практической деятельности.

Проблема исследования сознания селян в нэповских условиях перехода к рынку, выражающегося прежде всего через их отношение к реалиям социально-экономического и политического развития государства, чрезвычайно многогранна. Поскольку массовое сознание весьма многомерно, то воссоздание образующих его структур составляет главное условие его познания. Однако при всем обилии работ в отечественной историографии, в той или иной мере связанных с данной проблематикой, часть вопросов еще ждет своего решения. По-прежнему в поле зрения исследователей попадает лишь достаточно ограниченное число элементов сознания селян в условиях нэпа - их отношение к продналогу, к кооперации (преимущественно к ее производственным формам), к Советской власти и правящей партии, к проблемам социального расслоения. Ряд вопросов (например, выяснение мнений селян относительно возможности и необходимости регулирования социально-аграрной сферы в условиях перехода к рынку, эффективности новых форм организации труда на селе и т.д.) не ставится, или же априори имеют однозначную, если не конъюнктурную, трактовку. Что же касается всего "комплекса" этих составляющих в их взаимосвязи и взаимообусловленности, то подобные аспекты относятся к числу незатронутых. Другая проблема касается соотношения различных пластов сознания: традиционалистского (применительно к российскому крестьянству оно естественно имплицитно с издревле существовавшей общинной организацией земледельческого населения, основанной на дуализме - коллективной собственности на землю и частном характере производства и присвоения); условно говоря, индивидуалистского, начавшего вызревать во второй половине XIX века и получившего свое некоторое развитие после аграрной реформы П.А.Столыпина, и революционизированного. Думается, что вывод,

бытующий в литературе, о последовательном (хотя и не безболезненном) вытеснении первых двух пластов последним является во многом упрощенным. Каковы были способы "сосуществования" этих типов сознания в условиях нэпа? Что можно сказать о взаимопроницаемости и степени распространенности каждого из них в сельской среде? Существует ли связь между социальной структурой деревни 20-х гг. и указанными выше моделями? Если "да", то в чем её конкретные проявления? Иными словами, можно ли говорить о преимущественных носителях того или иного типа сознания или же обыденное сознание различных групп селян было синкретичным? Пока вопросов больше, чем ответов. Зияющие "пустоты" свидетельствуют: исследовательские успехи в указанной проблематике - только в поле системного подхода к имеющимся вопросам и к постановке новых.

Цель статьи достаточно конкретна - выявить глубину и характер взаимосвязей между элементами структуры сознания селян в первые годы нэпа, факторы ее формирования, а также типологизировать сознание жителей села на основе широкого спектра волнующих их проблем развития и государственного регулирования социально-аграрной сферы.

Источниковой базой исследования послужили письма и корреспонденции селян с ноября 1923 г. по июнь 1924 г. в редакции "Крестьянской газеты" и журнала "Крестьянка". Поскольку доля опубликованных писем ничтожно мала - 2-3%, - основной интерес представили достаточно хорошо сохранившиеся архивные материалы. В результате кропотливой эвристической работы были отобраны 376 писем, удачно репрезентирующих мнение селян по их социальным группам и регионам страны.

Большинство документов характеризуется наличием в них как информации "проблемно-содержательной", охватывающей комплекс суждений селян относительно социально-аграрной политики государства в условиях перехода к рынку, так и "анкетных" сведений (Ф., И., О. и адрес автора письма, его род занятий, пол и т.д.). "Анкетные" показатели формализовались методом "открытой" (номинальной) шкалы; ряду характеристик придавался альтернативный вид (например, партийность автора письма, его участие в гражданской войне, оформленность фабулы письма, определяемая посредством наличия или отсутствия заголовка и др.). При изучении же содержания писем совершенно необходимым оказалось обращение к контент-анализу полных текстов. В рассматриваемых нами письмах и корреспонденциях селян можно выделить несколько групп повторяющихся составляющих: 1. Суждения (например, "крестьянство есть круговая порука промеж собой, сосед без соседа никогда не проживет"); 2. Оценочные мнения по проблемам аграрного развития и его государственного регулирования ("...вот и деньги стали серебряные, и не теряют цену, - видно, что идет к лучшему"); 3. Требования ("...упорядочить единый сельхозналог, ...чтобы он соответствовал доходу хозяйства каждого крестьянина, чтобы избежать

недоимок..."); 4. Запросы и просьбы (запрос в редакцию о возможности и условиях расторжения договора между селянами и госработодателями ввиду невыполнения обязательств со стороны последних, просьба о содействии в перераспределении земель); 5. Жалобы (на самоуправство местных властей); 6. Предложения и пожелания ("хочется, чтобы было поменьше ломки старых укладов, которые испытаны целыми веками, или чтоб все эти ломки проходили с сознанием, дабы неприметно для самих крестьян"); 7. Предупреждения, обещания и заверения ("бойтесь темной неграмотной бедноты", "обещаем стремиться вперед быстрее; чем грамота скорее начинается, тем темнота скорей кончается").

Особую сложность процедуры контент-анализа писем и корреспонденций представляла неструктурированность текстов.

Возьмем в качестве примера небольшой по объему отрывок из письма в редакцию "Крестьянской газеты" "Кому густо, кому пусто" Н.Аржанова (Костромская губ.): "...крестьяне часто жалуются на тяжесть налогов, а если внимательно все рассмотреть и в самом деле - правда. Сравнить, например, советского служащего, который не так тяжело борется за свое существование, а получает каждый месяц большую ставку, которую крестьянину все трудовое лето не получить со своего хозяйства. Да вот еще что обидно. Вот, например, родился у нашего соседа советского служащего, или же умер ребенок, и за него он еще получает 15-25 рублей... А нет бы государству употребить эти деньги и роздать инвалидам и семьям погибших героев Красного Октября..." /1/. Уже на основе этих довольно кратких строк можно выделить по крайней мере 6 показателей, отражающих мировоззренческие установки, настроения, оценки автора письма, выраженные в тексте как в прямой форме, так и опосредованно: 1) констатация низкой доходности (рентабельности) единоличных хозяйств; 2) неудовлетворенность государственной системой налогообложения; 3) суждение об обнищании крестьянства; 4) констатация непомерно высоких ставок заработной платы работников среднего звена государственного аппарата; 5) мнение о становлении системы властных привилегий; 6) требование государственной программы социальной защиты крестьянства. Заметим, что первичные смысловые единицы, выявленные из большинства писем (как и из выше приведенного), не имеют явного терминологического выражения.

Количество показателей в комплексе 376 писем оказалось достаточно значительным - 192. Для облегчения последующей статистической обработки выделенных смысловых единиц последние были "укрупнены" в сюжетно-значимые темы (категории). К примеру, суждения о необходимости "перехода на многопольную систему обработки", констатации низкой продуктивности скота и низкого уровня агрокультуры, предложения о создании образцовых хозяйств в качестве примера землепользования, выражения неудовлетворенности ролью науки в сельскохозяйственном производстве, оценки высокой значимости помощи агронома при переходе к

многопольной системе обработки и т.д. можно объединить в одну категорию - "Осознание крестьянами необходимости улучшения сельскохозяйственной технологии и научно-практической помощи селу." Другой пример. В совокупности 192 первичных понятий довольно часто встречаются жалобы крестьян на высокие цены и качество продукции в частной торговле, выражения враждебного отношения к частным торговым посредникам, требования усиления налогового обложения кулаков и др. В результате агрегирования этих показателей возникла категория: "Неудовлетворенность крестьян оживлением частного предпринимательства с введением нэпа." Проанализировав таким образом все 192 показателя, в итоге нами было получено 84 категории. Затем в целях унификации массива для всех объектов (писем) по каждому признаку (категории) были определены частоты (встречаемость индикатора в тексте). Особенность БД в том, что конкретный признак (категория) имеет значения от 0 до 1,2..., т.е. он измерен по количественной шкале. Отсюда вытекают широкие потенции статистической обработки, в том числе продвинутыми метрическими методами. Отметим, что управление сформированной БД и вся обработка имеющейся информации осуществлялась при помощи программного обеспечения SPSS/PC+ - пакета статистических процедур для работы в области социальных наук (Statistical Package for Social Science). SPSS/PC+ предоставляет пользователю богатые возможности проведения анализа тех или иных групп признаков при решении как методических, так и конкретно-исторических задач. С помощью пакета SPSS/PC+ пользователь может легко манипулировать данными и методами, создавать новые переменные и модифицировать значения уже существующих, не прибегая к составлению программ. Предусмотрена возможность специального обращения со значениями, которые были пропущены. SPSS/PC+ содержит, помимо других, программы логлинейного анализа, алгоритмы множества мер расстояния и иерархических агломеративных методов кластерного анализа, несколько различных методов выделения факторов, их вращения и критериев, используемых при этом, и т.д.

Рассмотрим структуру сознания сельского населения в условиях перехода к рынку. Для облегчения восприятия выданной ЭВМ корреляционной матрицы последняя была представлена в графической форме посредством введения двух "порогов связи" ($r > 0,67$ и $0,4 < r < 0,67$) (см. рис.1).

Здесь просматриваются 4 основных комплекса связей. Отметим, что все полученные связи (по знаку) оказались прямыми. Первый блок тесно сопряженных между собой категорий раскрывает механизм функционирования "государственно-централистского" "пласта" в структуре сознания жителей села. Недоверчивое отношение крестьян к партийно-государственному ориентиру на смычку города с деревней (40) во многом дополнялось и усугублялось действиями диктатуры пролетариата, культивировавшей противоречия между рабочим классом и крестьянством, что четко

осознавалось последним (43). Отсюда, с одной стороны, - пренебрежительное отношение к земледельческому труду со стороны потребителей сельскохозяйственной продукции (44) и, как результат, миграция населения из села в город (47), нехватка трудовых ресурсов на селе (48); с другой, - выражение крестьянского недовольства непомерно высокими ставками заработной платы работников среднего звена государственного аппарата, констатация крестьянами процесса складывания системы властных привилегий (45). Квинтэссенцией понимания вышеозначенных тенденций развития являются требования крестьян широкомасштабной государственной программы социальной защиты (13) и правительственных дотаций непроизводственной сфере на селе (34). Как видим, характерной чертой данного "пласта" сознания выступает ориентация не на собственные силы, а на помощь "сверху", при этом обращение к "власть имущим" рассматривается как единственно возможное средство разрешения проблем. Могут возразить: низкая доходность единоличного хозяйства не позволяла обеспечить в должной мере социальные условия существования. Однако, думается, причины глубже, и связаны они с распространенностью среди определенной части крестьянства инфантилистских настроений. Ибо при наличии желания и стремлении к реальной самостоятельности выход есть. И видится он не столько в "хуторизации" крестьянских хозяйств, но и в их кооперации. Последняя не только "поднимает" своих членов на новый экономический уровень хозяйствования, но и опирается на крестьянство как источник для преодоления трудностей.

Второй комплекс, сердцевиной которого является осознание крестьянами необходимости развития кооперации, их убежденность в преимуществе коллективных форм ведения хозяйства над индивидуальными (1), констатация ими роста числа членов сельских кооперативов (2) и улучшения своего положения в результате развития кооперации (3), характеризует "традиционалистский" пласт сознания, применительно к российскому крестьянству связанный с общинной организацией земледельческого населения. Поясним сказанное. Издревле существовавшая сельская община представляла собой не только соседскую форму землепользования, но и выступала как трудовая кооперация и социальная общность. Кроме того, она имела уникальные психологические и правовые основы, обеспечившие ей завидную живучесть. Хотя после революции община формально перестала существовать, тем не менее в реальной жизни многие "мирские" порядки сохраняли свое значение. И именно исконный артельный дух крестьянского "мирского" сообщества побуждал ко множеству видов кооперации (коммуны, ТОЗы, кредитные, машинные, семенные товарищества и т.д. и т.п.). Поскольку один из вышеуказанных признаков фиксирует осознание крестьянами необходимости развития кооперации, отметим еще одну важную побудительную причину кооперативного движения, тесно сплетающуюся с историческим опытом

общины. В различных формах объединения (добровольных, естественно) крестьянин начала 20-х гг. видел самую надежную защиту от частного капитала, получившего свое оживление в связи с введением нэпа.

В то же время в данном комплексе сознания находила свое отражение и "командно-административная" тенденция регулирования аграрной сферы, выразившаяся, в частности, в стремлении форсированного вовлечения единоличных хозяйств в производственные кооперативы и совхозы, благодаря мощному государственному "покровительству" последних (67). Такая политика властей находила свою поддержку среди определенных слоев сельского населения, главным образом - пауперизированной и пролетаризированной его части. Интересно, что вступление крестьян на путь производственного кооперирования служило катализатором роста их культурно-образовательного уровня (64), роста общественно-политической активности женщин (83), доверия селян к Советской власти (78) и, естественно, укреплению авторитета партии на селе (69).

Отметим, объединение в данном комплексе взаимосвязей наибольшего числа категорий свидетельствует о высокой степени распространенности "традиционалистской" модели в структуре сознания сельского населения в нэповских условиях перехода к рынку.

Ключевым в понимании третьего комплекса является 81-й признак - стремление крестьян к росту зажиточности - причем, этот рост должен происходить на основе расширения со стороны государства свобод единоличного землепользования и товарооборота (79): "...необходимо поставить крестьянство на должную экономическую высоту... Пусть некоторые крестьяне хотя за это время и превратятся в кулачков, но черт с ними, ведь он каждый, чем больше кулак, тем больше он работник..." /2/. Это мнение крестьянина А.Шильцева органично перекликается с известными словами Н.И.Бухарина: "Крестьянам, всем крестьянам, надо сказать: обогащайтесь, развивайте свое хозяйство и не беспокойтесь, что вас прижмут. Мы должны добиться того, чтобы у нас беднота возможно быстрее исчезала, переставала быть беднотой." /3/. Важно, что и стремление крестьян к росту зажиточности, и поддержка ими государственной практики либерализации отношений землепользования и товарооборота тесно сопрягаются с их убежденностью в необходимости эволюционного перехода к нетрадиционным формам организации труда (80). Иными словами, допускается многообразие хозяйственных форм (более, придается этому немаловажное значение).

Итак, характерными чертами этого хозяйственно-психологического генотипа являются: установки на богатство (зажиточность), производственную эффективность, высокая ценность экономических свобод, отсутствие требований социальной защищенности хозяйств со стороны государства. Это позволяет обозначить его как "индивидуалистский", ориентированный на рыночные пути развития аграрной сферы. Безусловно, малое число категорий, образующих данный комплекс взаимосвязей,

позволяет говорить о нем лишь как о частном, а не как о доминантном (что, впрочем, закономерно, если учитывать российские традиции развития аграрной сферы, связанные с коллективными формами владения). Одновременно замкнутость линий данного блока убедительно свидетельствует: "индивидуалистский" "профиль" хозяйственно-психологического генотипа сельского населения в нэповских условиях перехода к рынку есть не произвольная мозаика, а некая целостность.

Четвертый комплекс взаимосвязей образуют категории, отражающие послевоенный упадок сельского хозяйства (36), технико-производственную слабость единоличных хозяйств (29), неудовлетворенность крестьян системой налогообложения (19), требования необходимости снижения налогов (20). Показательно наличие тесной связи между 20 и 22, 22 и 21 признаками и отсутствие таковой между 21 и 19. Это означает: требование крестьянами тщательной дифференциации хозяйств при их налогообложении не являлось свидетельством их неудовлетворенности этим налогообложением (применительно к собственному хозяйству), а выдвигалось исключительно с целью того, чтобы "...побольше бы богач дал своих излишков государству." Такое "нездоровое" отношение к чужим доходам в совокупности с отсутствием в данном комплексе признака, отражающего необходимость повышения доходности единоличного хозяйства (46) позволяет, на наш взгляд, говорить о "люмпенско-пролетаризированной" направленности этого пласта в структуре сознания жителей села. Убедительным подтверждением тому является связь между указанным 22 и 4 (расслоение крестьянства, пролетаризация его части) признаками.

Выяснить причины, детерминировавшие вышерассмотренную структуру сознания селян, позволил факторный анализ. На основе метода главных компонент и критерия Кайзера было получено 10 значимых факторов, объяснивших 63,3% дисперсии исследовавшихся 26 ключевых категорий, значения частот встречаемости которых в документах являются наибольшими. I фактор - "Пропаганда преимуществ организации быта, досуга и труда крестьян в коммунах и совхозах." II - "Застой сельского хозяйства, усугубляемый системой налогообложения. Надежды на его преодоление благодаря регулирующей функции государства." III - "Антикрестьянская направленность социальной политики диктатуры пролетариата. Обострение выше обычного противоречий между городом и селом." IV - "Послереволюционный вакуум культуры, радикализирующий рост духовных потребностей крестьян." V - "Меры наибольшего благоприятствования властей кооперативным формам организации сельскохозяйственного производства." VI - "Проникновение идей технологического прогресса в аграрную сферу." VII - "Пролетаризация крестьянства. Усиление иждивенческой психологии, люмпенизация сознания." VIII - "Активизация частного предпринимательства в аграрной сфере в связи с новой экономической политикой государства.

Неудовлетворенность крестьян государственными мерами защиты их хозяйств. Стремление крестьянства к простым формам кооперирования в целях противостояния разрушающей роли частного капитала." IX – "Усиление централизации государственного руководства экономикой." X – "Традиции крестьянской веры в справедливость действий высшей власти в стране."

Таким образом, действие лишь двух (VI-го и VIII-го) из 10-ти выделенных факторов тем или иным образом было связано с рыночными механизмами функционирования аграрной сферы. Их доля составляет 10,1% (или чуть более 1/6) в суммарной дисперсии. Такая ситуация являлась отражением незавершенности самой концепции нэпа с сохранением в ней множества противоречий, главное из которых – между "рыночным" нэпом (или же экономической целесообразностью) и "ортодоксальными" представлениями марксизма о переходе к социализму и о самом социализме как "планомерном, бестоварном" обществе с концентрацией основных средств производства в руках государства, экспроприацией земельной собственности и устранением различий между городом и деревней. Наличие в модели III, V, IX факторов подтверждает это положение.

Анализ факторной модели позволяет сделать вывод и о реальной многовариантности развития сознания исследуемой социальной общности /4/. В этой связи выглядят легковесными и необоснованными то и дело возникающие в современной публицистике попытки "приписывания" какой-либо единой ментальности всему крестьянству. Особенно характерно в этом отношении мнение И.Клямкина, что "...наш крестьянин в силу своего "добржуазного" характера предпочел коллективизацию кулаку, ...готов был записать и часто записывал в кулаки крепкого середняка, ...благословил удаление из деревни наиболее зажиточной и приспособленной к свободному рыночному хозяйству, наиболее "буржуазной" части населения" /5/. Такого рода грубо-уравнительные, левозэкстремистские настроения, безусловно, были, однако, они являли собой лишь одну из возможных (но не единственную и доминирующую) тенденций развития сознания сельского населения в 20-е гг. В этом плане вполне актуальным представляется вопрос об основных типах сознания жителей села начала 20-х гг. Иначе – какие можно выделить группы селян с точки зрения волнующих их проблем развития и государственного регулирования социально-аграрной сферы в нэповских условиях перехода к рынку, каков был удельный вес каждой из них?

Типологизацию сознания исследуемой социальной общности на основе писем их представителей корректно можно провести с помощью кластер-анализа. В обработку были включены все 84 категории текстов писем. Агломеративно-иерархическая кластеризация осуществлялась на основе метода полной связи – максимальное локальное расстояние групп объектов (COMPLETE); при этом мерой близости объектов служило "манхэттенское" расстояние (BLOCK). Нами экспериментально установлено: сочетание

COMPLETE-метода и BLOCK-расстояния, содержащихся в SPSS/PC+, наиболее перспективно для анализа множества объектов, описываемого достаточно дискретными признаками, имеющими "хи-квадрат" распределение. Последними и характеризуем наш массив данных, впрочем, как и большинство формализованных в результате контент-анализа текстов частных писем (обращений).

Было выделено 24 кластера (см.рис.2), объединяющихся в 4 макротипа массового сознания.

Численно наибольшим предстает макротип **А** (59,2% от общего числа писем в редакцию). Эта группа, в отличие от остальных, характеризуется наличием в ней представителей всех социальных категорий сельского населения: 57,1% являлись крестьянами-единоличниками, 15,6% - членами производственных кооперативов, 2,6% - кустарями и ремесленниками, 2,6% - рабочими совхозов и госпредприятий, 6,5% - батраками, наемными работниками частных предприятий, 1,3% - предпринимателями, 1,3% - председателями производственных кооперативов; освобожденные советские, партийные, комсомольские, профсоюзные работники, а также служащие и интеллигенция составляли по 6,5%. Что же объединяло такую довольно "разноликую" массу сельских жителей? Анализ показал, что данную группу селян в первую очередь волновали проблемы роста национализма, притеснения русскоязычных в национальных районах, роста цен на местном рынке, неэффективности производства коммун и совхозов, вопросы сохранения общинных традиций. Кроме того, эти представители села выражали свою глубокую обеспокоенность такими проблемами как хищническое отношение человека к природе, обострение противоречий между мужчинами и женщинами в распределении сфер трудовой деятельности, игнорирование мнений крестьянства при решении вопросов села, отсутствие согласованности в действиях местных органов власти, самоспаивание деревни, пьянство должностных лиц. Итак, жители села этой группы - люди, фиксирующие взрыв традиционных устоев деревни (причем, этот "взрыв" вызван не только (и не столько) новой экономической политикой, но и революцией, гражданской войной; некоторые же проблемы уходят своими корнями еще в дореволюционные времена); осознавшие "болезнь", кризис "социально-экологической ниши" сельского общества. Однако, констатируя катастрофическую ситуацию в аграрной сфере и осознавая необходимость ее изменения, "социальные экологи", как можно условно обозначить жителей села данной группы, твердо убеждены в том, что все преобразования должны осуществляться не радикально-революционными средствами, а постепенным, эволюционным путем. Показательно отсутствие в данном спектре приоритетно волнующих проблем, вопросов, касающихся государственного регулирования социально-аграрной сферы. Именно этим как раз и объясняется "всеядность" анализируемого нами макротипа с точки зрения "анкетных" характеристик лиц, его составляющих. Ибо, в отличие от вопросов

государственного вмешательства в экономику, неизбежно поляризующих людей, проблемы социальной экологии могут объединять (и объединяют, как видим, на самом деле) достаточно пеструю по своему составу массу селян - мужчин и женщин, партийных и беспартийных, селькоров и не принадлежащих к их числу, крестьян-единоличников и рабочих совхоза, батраков и интеллигенцию и т.д.

Жители села, объединенные в макротип **Б** (13,1% от общего числа авторов писем в редакцию), констатируя послевоенный упадок сельского хозяйства, технико-производственную слабость единоличных хозяйств, отсутствие паритета цен на промышленные товары и сельскохозяйственную продукцию, низкую доходность единоличных хозяйств и осознавая необходимость её повышения, одновременно обладают глубокой верой в финансовую стабилизацию экономики аграрной сферы и улучшение материального положения лиц, занятых в ней. Причем рост собственной покупательной способности связывается ими прежде всего с:

- развитием контрактов между госорганами и кооперацией;
- развитием промыслов и государственной практики расширения свобод единоличного землепользования и товарооборота;
- разумной ценообразовательной и денежной (в том числе кредитной) политикой государства;
- установлением "рыночной" системы налогообложения, ориентированной не на максимальное изъятие средств из производства, а на свободу хозяйствования, рациональное капиталовложение и, в итоге, на рост экономической эффективности.

Стало быть, жители села данной группы, определяемые нами как *сторонники социально ориентированной рыночной экономики*, понимают необходимость регулирования ее со стороны государства. Однако, последнее ассоциируется ими не с административно-директивным государственным вмешательством, а затрагивает, в первую очередь, финансово-кредитные рычаги, сферу налоговых правил и законодательных актов.

Примечательно, что среди сторонников гибкого государственного (посредством экономических методов) регулирования аграрной сферы доля индивидуальных хозяев предстает максимальной: 82,4% от общего числа селян группы **Б** являлись крестьянами-единоличниками, 6% - кустарями и ремесленниками (что больше, чем во всех остальных типах), тогда как рабочих совхоза, членов сельскохозяйственных производственных кооперативов (и, естественно, их председателей) в ней нет. Следовательно, хотя кооперация и возникла исторически как компонент рыночного хозяйства, но вследствие проводившейся государством диктатуры пролетариата политика протекционизма путем создания режима наибольшего благоприятствования её производственным формам - ТОЗам, артелям, коммунам, - члены последних, так сказать "не ощущали" самого "духа" рыночной стихии.

Селяне, вошедшие в макротип **В** (13,1% от общего числа авторов писем), в числе первостепенных по значимости проблем развития и мер государственного регулирования социально-аграрной сферы выдвигают следующие:

- требование увеличения капитальных вложений и финансовых дотаций в аграрную сферу;
- требование правительственной программы социальной защиты крестьянства;
- недовольство отсутствием системы госстрахования на селе;
- требование централизованных государственных дотаций в непроизводственную сферу на селе;
- требование беднейших усиления государственного рентно-налогового обложения зажиточных.

Очевидно, все вышеуказанные черты сознания характеризуют его носителей как приверженцев активного (и административного в том числе) вмешательства государства в экономику, как сторонников усиления централизованного руководства социально-аграрной сферой. Заметим, что среди ведущих категорий удельный вес требований в макротипе **В** является наибольшим (по сравнению с другими группами). Это закономерно, ибо отказ от личной ответственности в результатах производства, сопровождающийся распространением настроений уравнилельности и иждивенчества, неизбежно порождает особую, можно сказать, агрессивность и ожесточенность требований к партийно-государственной власти. В случае же обмана ожиданий, виновник легко находится в числе обещавших "безбедное существование". Отсюда - вполне понятно недоверие селян группы **В** к аграрной политике Советской власти: пообещав, образно говоря, "светлое будущее", наступление которого казалось таким близким после победы социалистической революции, рабоче-крестьянская власть с переходом к нэпу фактически способствовала распространению того, против чего, согласно укоренившемуся в сознании мнению, и боролась - социального неравенства. В этой связи неудивительно наличие в спектре проблем, приоритетно волнующих данную группу селян, таких как расслоение крестьянства, пролетаризация его части; обнищание, низкая покупательная способность селян; недовольство беднейших сохранением имущественного ценза в системе школьного образования; неудовлетворенность системой налогообложения в плане желаемости усиления давления налогового пресса на крепнущие и зажиточные слои деревни. Учитывая социальный состав жителей села анализируемой группы - 88% их числа составляли крестьяне-единоличники - можно достаточно определенно говорить, что подавляющее большинство из них являли собой деклассированные, люмпен-пролетарские элементы деревенского населения. Остальные же 12% являлись мелкими советскими служащими (что больше, чем в других кластерах). Как видим, представителей этих двух социальных категорий сельского населения объединила в один макротип

общая, так сказать, "централистская" ориентированность их сознания, подразумевающая активное вмешательство государства как в производственную, так и в непроизводственную сферы аграрного сектора.

Наиболее обособленным от всех других групп предстает макротип **Г** (14,6% от общего числа авторов писем в редакцию). Из рис.2 видно, что максимально сходными являются макротипы **А** и **Б** (и действительно, "социальные экологисты" и сторонники регулируемой (посредством "гибких" государственных мер) рыночной экономики близки между собой в плане неприятия радикально-революционных, административно-командных средств разрешения проблем); затем к ним примыкает макротип **В**, и на заключительном этапе - **Г**. Однако из сказанного вовсе не следует вывод о незначительной роли последнего макротипа в палитре массового сознания селян в нэповских условиях перехода к рынку. Напротив, можно говорить о существенной значимости данного типа, убедительным доказательством чему является заметное "смещение" центра условного (возможного) объединения селян с точки зрения "общности" черт их сознания вправо, в сторону макротипа **Г**.

С точки зрения социального состава селян группы **Г** подавляющее их большинство (57,9%) являлись членами производственных кооперативов и рабочими совхозов; 36,8% относились к категории освобожденных советских, партийных, комсомольских работников и к интеллигенции, и лишь 5,3% были представлены крестьянами-единоличниками. Кроме того, удельный вес коммунистов (комсомольцев) в этом макротипе предстает наибольшим, практически каждый третий являлся членом РКП(б) или РКСМ. Налицо - заметное "покраснение" состава лиц данной группы по сравнению с другими. Анализ приоритетно волнующих проблем показывает, что все они так или иначе связаны с развитием социалистического уклада в аграрной сфере.

Отмечая соблюдение принципов выборности руководства и самоуправления в коммунах, рост общественно-политической активности женщин с переходом к новым формам организации труда, жители села группы **Г** твердо убеждены в преимуществе коллективных форм ведения хозяйства (по сравнению с индивидуальными), ибо последние, по их мнению, открывают невиданные возможности для самореализации духовных и трудовых потребностей крестьянства, а также способствуют значительному улучшению его материального положения. Отсюда - рост доверия к Советской власти и укрепление авторитета партии на селе. Причем, если последнее жителями села, вошедшими в первые два кластера макротипа **Г** (21 и 22) связывается с организационно-производственной стороной существования социалистических форм земледелия, то авторами писем, образующими 23 и 24 кластеры, - с культурно-просветительными аспектами развития колхозов и совхозов.

Из рис.2 видно, что наиболее "близким" к макротипу **Г** оказался макротип **В**. В самом деле, приверженцев активного (в том числе и

административного) государственного вмешательства в экономику и "активистов социалистических форм земледелия" (как условно обозначены нами селяне группы Г) сближала их общая ориентированность на помощь "сверху", со стороны партийно-государственных структур. Разница лишь в том, что если в первом случае такая "поддержка" рассматривалась лишь как желаемость, то во втором она имела черты реальной политики.

----- $r > 0,67$

----- $0,4 < r < 0,67$

Рис.1. СТРУКТУРА СОЗНАНИЯ СЕЛЬСКОГО НАСЕЛЕНИЯ
В НЭПОВСКИХ УСЛОВИЯХ ПЕРЕХОДА К РЫНКУ

Номера и названия признаков: **1** - Осознание крестьянами необходимости развития кооперации, их убежденность в преимуществе коллективных форм ведения хозяйства над индивидуальными; **2** - Рост числа членов сельских кооперативов; **3** - Улучшение положения крестьянства в результате развития кооперации; **4** - Расслоение крестьянства, пролетаризация его части; **5** - Скрытые формы эксплуатации лиц наемного труда кулаками; **6** - Рост национализма, притеснения русскоязычных в национальных районах; **8** - Рост цен на местном рынке; **9** - Рыночная конкуренция как условие снижения цен; **13** - Требование правительственной программы социальной защиты крестьянства; **14** - Удовлетворенность крестьянства государственной денежной и ценообразовательной политикой в аграрной сфере; **15** - Удовлетворенность селян деятельностью потребкооперации; **16** - Развитие договорных отношений между госорганами и кооперацией; **19** - Неудовлетворенность индивидуальных хозяев системой налогообложения; **20** - Необходимость снижения налогов; **21** - Требование крестьянами тщательной дифференциации хозяйств при их налогообложении; **22** - Требование беднейших усиления государственного рентно-налогового обложения зажиточных (кулаков); **23** - Желание бедняками правительственной линии к уравнильно-трудовому распределению земли; **28** - Невнимание местных органов власти к проблемам крестьянства; **29** - Техничко-производственная слабость единоличных хозяйств; **30** - Несоответствие материально-технической базы культурно-просветительных и образовательных учреждений обострившимся запросам селян; **32** - Отсутствие квалифицированных учительских кадров на селе; **34** - Необходимость централизованных государственных дотаций непромышленной сфере на селе; **35** - Незнание крестьянством законов и кодексов, регулирующих аграрные отношения; **36** - Послевоенный упадок сельского хозяйства; **37** - Промыслы как параллельный доход крестьянина; **39** - Стремление частных предпринимателей к получению госзаказов; **40** - Недоверчивое отношение крестьян к партийно-государственному ориентиру на смычку города с деревней, их убежденность в противоречиях городского и сельского образов жизни; **43** - Осознание крестьянами государственно культивируемых противоречий между собой и рабочим классом; **44** - Пренебрежительное отношение к земледельческому труду со стороны потребителей сельскохозяйственной продукции; **45** - Непомерно высокие ставки заработной платы работников среднего звена государственного аппарата; становление системы властных привилегий; **47** - Миграция населения из села в город; **48** - Нехватка трудовых ресурсов на селе; **49** - Непосильный физический труд детей и стариков; **52** - Недоверие крестьян к аграрной политике Советской власти; **53** - Отход крестьян на заработки в город; **54** - Невыполнение госработодателями договорных обязательств по отношению к селянам; **55** - Неудовлетворенность крестьян низкими размерами оплаты труда в госпредприятиях (хозяйствах); **56** -

Отсутствие поддержки членов коммун местными крестьянскими комитетами взаимопомощи; **57** - Недовольство лиц наемного труда (батраки, совхозные рабочие) степенью помощи "Всеработземлеса"; **58** - Признание лицами наемного труда "Всеработземлеса" организационной формой защиты своих интересов; **60** - Недовольство крестьян сохранением имущественного ценза в системе школьного образования; **62** - Осознание селянами необходимости усиления политико-просветительной и партийной работы в деревне; **63** - Хищническое, потребительское отношение к природе; **64** - Рост культурно-образовательного уровня селян; **65** - Рост покупательной способности крестьян; **66** - Укрепление материально-технической базы культурного развития села; **67** - Особое "покровительство" местных советских и партийных органов сельскохозяйственным производственным кооперативам в противовес единоличным крестьянским хозяйствам; **68** - Укрепление материально-технической базы сельскохозяйственных производственных кооперативов; **69** - Укрепление авторитета партии на селе; **70** - Классовый принцип организационного строения сельскохозяйственных производственных кооперативов; **71** - Возможность самореализации крестьянами духовных и трудовых потребностей в коммунах и совхозах; **72** - Удовлетворенность совхозных рабочих социальной политикой государства; **73** - Противоречие коммунистической идеологии и традиций общины; **74** - Становление системы партийно-политического просвещения на селе; **77** - Убежденность в определяющем значении общественного мнения односельчан на все стороны крестьянской жизни; **78** - Рост доверия селян к Советской власти; **79** - Поддержка крестьянами государственной практики расширения свобод единоличного землепользования и товарооборота; **80** - Убежденность крестьянства в необходимости эволюционного перехода к новым формам организации труда; **81** - Стремление крестьян к росту зажиточности; **82** - Неэффективность производства коммун и совхозов; **83** - Рост общественно-политической активности женщин; **84** - Соблюдение принципов выборности руководства и самоуправления в жизни коммун; **85** - Пол автора письма: женский - 1, мужской - 2; **88** - Партийность автора: член РКП(б) или РКСМ - 1, беспартийный - 0; **90** - Социальное положение автора: частные производители - 1, представители обобществленного сектора - 0.

ПРИМЕЧАНИЯ

1. РГАЭ, ф.396, оп.1, д.4, л.2.
2. РГАЭ, ф.396, оп.2, д.23, л.51.
3. Правда. 1925. 24 апреля.
4. Основанием для такого вывода служат сравнительно небольшая амплитуда колебаний значений вкладов каждого из 10-ти выделенных факторов в суммарную дисперсию, а также отсутствие в модели одного-двух "всеядных" доминирующих факторов.

5. Клямкин И. Какая улица ведет к храму //Новый мир. 1987, N.11, с.135.

ПРОГРАММА FUZZYCLASS: НОВЫЕ ВОЗМОЖНОСТИ АРХЕОЛОГИЧЕСКОГО
ИССЛЕДОВАНИЯ

(Типология поселений Алтая раннего железного века)

М.Т.Абдулганеев, В.Н.Владимиров

Проблема классификации и типологии археологических памятников неизбежно встает перед исследователем-археологом, выходящим на уровень исторической интерпретации материала, рассматривающим вещественные источники как отражение реальной жизни людей во всех ее аспектах. Надежным средством для этого являются методы автоматической классификации /1/. Однако традиционная классификация предполагает жесткое отнесение объекта к одному и только к одному классу /2/. В противовес этому появившееся в последние десятилетия понятие "размытая", "нечеткая" (fuzzy) классификация дает возможность отнесения одного объекта сразу к нескольким классам с определением веса принадлежности к каждому из них, что во многих случаях более адекватно исторической действительности, а также позволяет выделять ядро и периферию каждого класса.

Опыт применения такого подхода показывает его большую плодотворность /3/. В настоящее время он реализован в ряде статистических пакетов для персональных компьютеров /4/. Нами использовалась программа FUZZYCLASS, разработанная в Лаборатории исторической информатики Московского университета /5/. Выражаем глубокую благодарность Л.И.Бородкину и И.М.Гарсковой за возможность работы с оригинальным программным продуктом.

Целью нашего исследования, результаты которого частично опубликованы /6/, является классификация поселений раннего железного века степного, лесостепного и предгорного Алтая, определение основных культурных компонентов оставившего их населения, уточнение и дополнение существующей хронологической схемы.

Эпоха раннего железного века на Алтае представлена большим количеством археологических памятников: курганных и грунтовых могильников, поселений, культовых мест. В середине 80-х годов стало ясно, что традиционные схемы исторического развития, созданные в 50-е гг. /7/, не отражают в полной мере того разнообразия, которое предстает перед глазами археолога. Интенсивные миграции и контакты населения привели к появлению большого числа различных культурных традиций и в то же время способствовали возникновению переходных форм. В этих условиях очень трудно разобраться, где "чистые", а где смешанные памятники, какие группы населения являются автохтонными, а

какие пришлыми, и вообще, сколько же разных групп населения представлено в археологическом материале, какие это группы, что означают различия между ними. В то же время материал поселений сильно отличается от материала могильников, и не всегда понятно, как провести их корректное сопоставление. В результате исследователями созданы различные культурно-исторические схемы развития, часто сильно отличающиеся друг от друга /8/.

Наше внимание привлекли поселения рассматриваемого времени. Достаточно большое их количество, весьма существенные отличия друг от друга, определенная разница в датировке и в то же время безусловная принадлежность к одному историко-культурному пласту побуждают к попыткам структурировать материал, исходя из каких-то признаков, которые являются едиными для всех памятников и в то же время различают их.

Традиционная археология всегда придавала большое значение орнаментации керамики, считая ее мощным и в то же время чувствительным культурным индикатором. Поэтому вполне естественным выглядит выбор видов орнамента в качестве признаков для анализа. Проанализированный материал позволил выделить 12 таких видов, которые в той или иной степени присущи всем или абсолютному большинству памятников, и в то же время представлены на разных поселениях в разном количестве (рис.1):

1. Наколы различной формы в один ряд.
2. Жемчужник в один ряд.
3. Жемчужник с разделителем (ямки, насечки, резные линии, уголок палочки и т.д.).
4. Жемчужник, разделенный оттисками гребенчатого штампа.
5. Наколы в сочетании с резными линиями, поставленными наклонно или образующими елочку или зигзаг.
6. Оттиски гребенчатого штампа, поставленные наклонно в один ряд или образующие елочку.
7. Наколы различной формы в 2-4 ряда.
8. Жемчужник с разделителем в сочетании с рядами наколов сверху или снизу либо сверху и снизу одновременно.
9. Жемчужник или жемчужник с разделителем в сочетании с резной елочкой.
10. Жемчужник с разделителем в сочетании с резными наклонными линиями или сеткой по верху.
11. Резные наклонные линии или резная елочка.
12. Валик или валик в сочетании с жемчужником, жемчужником с разделителем или наколами, или зажимы.

Для всех 39 представленных объектов был осуществлен подсчет каждого вида орнамента в процентах. В результате получилась матрица исходных данных /9/, которая далее может анализироваться различными методами.

Рассматриваемые поселения территориально находятся на сравнительно небольшом расстоянии друг от друга, хотя располагаются в различных природных условиях: лесостепи, степи и в предгорьях. А priori представляется, что это разделение имеет не только географический, но и определенный культурный смысл. Вместе с тем отдельные группы населения, представленные изучаемыми памятниками, отнюдь нельзя назвать изолятами. Достаточно интенсивные миграционные процессы в эпоху раннего железа на территории Алтая приводили к определенному смешению населения. На это указывает и предварительный анализ материала, свидетельствующий, что три основных орнамента (1,2,3) представлены на всех памятниках.

Методика работы заключалась в автоматической классификации поселений с помощью программы нечеткой классификации FUZZYCLASS. Поскольку программа позволяет задавать число классов, мы последовательно увеличивали его, начиная с двух и доведя до десяти. Анализ распределения памятников по классам привел нас к выводу, что наиболее оптимальным является выделение пяти групп. При числе классов более пяти картина классификации начинает, на наш взгляд, становиться слишком дробной и фактически размываться.

Обращает на себя внимание то обстоятельство, что классификация памятников тесно связана с их географическим положением. Исходя из этого выделенные нами группы памятников были названы в соответствии с теми ландшафтными зонами, где они локализируются:

- лесостепная группа (11 поселений);
- предгорная группа 1 (7 поселений);
- степная (4 поселения);
- предгорная группа 2 (3 поселения);
- смешанная группа (5 поселений).

По мере постепенного увеличения числа классов сначала выделяются лесостепная и предгорная 1 группы как самые многочисленные и монолитные, затем к ним последовательно добавляются степная, смешанная и предгорная 2. Поскольку программа показывает вес принадлежности каждого памятника к каждому классу, 9 памятников из 39 (23%) оказались не принадлежащими к какому-либо одному классу даже на 50%.

Для того чтобы четко представлять себе отличие групп друг от друга, необходимо проанализировать распределение признаков-орнаментов в каждой группе. Но, учитывая достаточную культурно-хронологическую близость всех групп и рассчитывая показать различие между ними более рельефно, мы решили выделить в каждой группе ядро из памятников на основе веса их принадлежности к каждой группе не менее 67%, т.е. двух третей. Таких объектов в перечисленных выше группах оказалось соответственно 9, 6, 3, 3, 4. Таким образом отсеялось еще 5 памятников, а объектов, которые можно назвать эталонными для своих классов, осталось 25 или 64% от исходной совокупности.

Анализ орнаментации в группах велся на основе подсчетов коэффициентов вариации в каждой группе эталонов по каждому признаку и сравнения этих результатов с общей статистикой орнаментации по всей совокупности, в которой лишь три вида орнамента (1, 2, 3) имели коэффициент вариации меньше 100% (соответственно 90,9%, 62,3% и 58,1%). В группах нами учитывались в качестве характерных признаков лишь те орнаменты, которые имели коэффициент вариации менее 50%. Практически полностью отсутствуют орнаменты 5 и особенно 7. Вероятнее всего, наличие этих орнаментов на керамике эпохи раннего железа следует считать поздним признаком - влиянием населения кулайской или майминской культур.

В ядро первой группы вошло 9 памятников лесостепного Приобья: поселения Новоалтайское, Фирсово 4, 10, 15, Малый Гоньбинский Кордон 1, Костенкова Избушка, Ольгинка 1, Васино 2, Лесная Пасека 1 (рис.2-1). Расположены они главным образом по правому берегу р.Оби, за исключением двух, находящихся значительно южнее остальных. Отличительными орнаментами для этой группы являются 1 ($V=27,6\%$) и 3 ($V=20,9\%$). Мы интерпретируем ее как группу поселений, имеющую орнаментацию, характерную для автохтонного населения, и сопоставимую с могильниками староалейского типа (конец VI - II в. до н.э.).

Шесть памятников северных предгорий Алтая - Точильное 1, Енисейское 1, Малоугренево, Майма 3, Стан-Бехтемир, Степь-Чумыш (рис.2-4) - составили ядро предгорной группы 1. Их объединяют орнаменты 1 ($V=35,1\%$), 2 ($V=33,1\%$), 3 ($V=25,7\%$) и 12 ($V=38,2\%$). Последний орнамент свидетельствует об участии в формировании культуры пазырыкского населения Горного Алтая. Что касается орнамента 1, то он, вероятнее всего, указывает на влияние населения Юго-Западного Алтая. Группу можно соотнести с основной массой могильников быстрианской культуры и датировать в пределах конца VI - II в. до н.э.

В ядро следующей группы вошло 3 поселения, находящихся в контактной зоне лесостепи с Кулундинской степью: Быково 3, Киприно, Островное 3 (рис.2-2). Отличительными орнаментами являются 1 ($V=9,4\%$), 2 ($V=14,4\%$), 4 ($V=28,2\%$), 11 ($V=44,5\%$). Характерным представляется и отсутствие орнамента 10. Мы склонны интерпретировать полученные данные как участие в формировании этой группы, с одной стороны, населения саргатской культуры, с другой - слабо изученной пока культуры раннего железного века Юго-Западного Алтая. По внешнему облику керамика этой группы очень близка к новосибирскому варианту большереченской культуры и датируется, вероятно, концом VI - V в. до н.э. Поселения данной группы следует соотносить с наиболее ранними могильниками каменной культуры.

Памятники северных предгорий Алтая - Солонцы 1, Новозыково 3, Березовка 2 (рис.2-3) - составили ядро предгорной группы 2. Они отличаются от других большой долей в орнаментации орнамента 8

(V=8,7%). С нашей точки зрения эти поселения являются крайними в хронологической шкале быстрянской культуры. Отсутствие в их орнаментации признаков, характерных для последующих культур (кулайская, майминская), а также определенное сходство с бийской керамикой заставляют предполагать их возраст достаточно ранним (конец VI - V в. до н.э.).

В ядро последней группы вошло три поселения приобской лесостепи и одно предгорное: Фирсово 6, Ближние Елбаны 12, Турина Гора 1 и Майма 1 (рис.2-5). Явное преобладание орнамента 2 (V=12,7%) свидетельствует о том, что на этих памятниках в наиболее чистом виде проявилось влияние населения Юго-Западного Алтая. Самая вероятная датировка этой группы - конец VI - V в. до н.э.

Выделенные нами группы мы считаем возможным трактовать как особые типы поселений Алтая в эпоху раннего железного века, отражающие различия в орнаментации в результате существования и смешения как минимум трех массивов населения: автохтонного, горного и степного. Выделение ядер групп позволило определить три наиболее ранние группы поселений и две группы основного существования культур. Оказалось возможным вычлнить ранние и поздние признаки, а также влияние соседнего населения. В дальнейшем при сопоставлении с соответствующими группами могильников возможно выявление более четкой картины культурно-исторического развития населения Алтая середины и второй половины I тыс. до н.э.

С нашей точки зрения, программа FUZZYCLASS дает огромные возможности для археологической классификации, особенно в тех весьма частых случаях, где бывает очень трудно четко отнести тот или иной объект к определенному классу. Без сомнения, она должна занять достойное место в историко-типологических исследованиях.

ПРИМЕЧАНИЯ

1. Количественные методы в исторических исследованиях. М., 1984, с.268-276; Бородкин Л.И. Многомерный статистический анализ в исторических исследованиях. М., 1986, с.188; Федоров-Давыдов Г.А. Статистические методы в археологии. М., 1987, с.180-189; Мандель И.Д. Кластерный анализ. М., 1988, с.176; Факторный, дискриминантный и кластерный анализ. М., 1989, с.139-214; Doran J.E., Hodson F.R. Mathematics and computers in archeology. Edinburgh, 1975, p.158-186.

2. Ср.: "...Полученные классы должны исключать друг друга..." (Тульчинский Г.Л., Светлов В.А. Логико-семантические основания классификации // Типы в культуре. Л., 1979, с.22.)

3. Ковальченко И.Д., Бородкин Л.И. Вероятностная многомерная классификация в исторических исследованиях // Математические методы и ЭВМ в исторических исследованиях. М., 1985, с.6-30; Бородкин Л.И.

Многомерный статистический анализ в исторических исследованиях, с.94-99.

4. См. например, Осокина Е., Сатаров Г. Прикладные возможности статистического пакета "CLAMS". Анализ распределительной политики советского правительства в период карточной системы 1928/29 - 1935 гг. // История и компьютер: Новые информационные технологии в исторических исследованиях и образовании. St.Katharinen, 1993, с.77-88.

5. Бородкин Л., Гарскова И. FUZZYCLASS: гибкое программное средство построения многомерной типологии объектов социальной природы // Информационный бюллетень комиссии по применению количественных методов и ЭВМ в исторических исследованиях при отделении истории РАН и ассоциации "История и компьютер". - М., 1992, N.7, декабрь, с.51-53; Бородкин Л., Гарскова И. Программное обеспечение FUZZYCLASS в историко-типологическом исследовании // История и компьютер: Новые информационные технологии в исторических исследованиях и образовании. St.Katharinen, 1993, с.89-104; Garskova I., Borodkin L. FUZZYCLASS: A New Tool for Tipological Analysis in Historical Research // History and Computing in Eastern Europe. St.Katharinen, 1993, p.71-86.

6. Абдулганеев М.Т., Владимиров В.Н. Статистическая обработка керамики поселений раннего железного века лесостепного и предгорного Алтая // Проблемы хронологии и периодизации археологических памятников Южной Сибири. Барнаул, 1991, с.115-117; Абдулганеев М.Т., Владимиров В.Н. Статистическая обработка керамики и периодизация поселений лесостепного и предгорного Алтая второй половины I тыс. до н.э. // Проблемы хронологии в археологии и истории. Барнаул, 1991, с.93-107; Владимиров В.Н., Абдулганеев М.Т. К методике многомерного статистического анализа орнаментации керамики поселенческих комплексов // Теория и методика исследований археологических памятников лесостепной зоны. Липецк, 1992, с.36-38; Vladimirov V.N. Multivariate Classification of the Early Iron Era Altai Settlements: Computer Based Analysis of Ceramics // The Art of Communication: VIII International ANS-Conference. Graz, Austria, August 24-27, 1993. Vienna, 1993, p.90-91; Vladimirov V.N. New Information Technologies in Historical Research and Education: The Siberian Experience // History and Computing in Eastern Europe. St.Katharinen, 1993, p.31-38.

7. Грязнов М.П. История древних племен Верхней Оби по раскопкам близ с.Большая Речка // МИА 48. М.; Л., 1956, с.265

8. Алехин Ю.П. Лесостепной Алтай в скифское время (V-I вв. до н.э.) // Скифская эпоха Алтая. Барнаул, 1986, с.36-38; Уманский А.П. О культурной и этнической принадлежности курганов раннего железного века в лесостепном Алтае // Барнаулу 250 лет. Барнаул, 1980, с.50-53; Могильников В.А. О культурах западносибирской лесостепи раннего железного века (итоги и проблемы изучения) // Скифо-сибирское культурно-историческое единство. Кемерово, 1980, с.41-50; Суразаков

А.С. Горный Алтай и его северные предгорья в конце VI - начале II вв. до н.э. Автореферат диссертации ... канд.ист.наук. Новосибирск, 1985, с. 22; Киреев С.М. Работы на Майминском комплексе в 1990-91 гг. // Проблемы сохранения, использования и изучения памятников археологии. Горно-Алтайск, 1992, с.55-56.

9. Не приводится из-за ограниченного объема статьи.

К ВОПРОСУ ОБ АВТОРСТВЕ ПРЕДСМЕРТНОГО ПИСЬМА Б.В.САВИНКОВА
- ОПЫТ КОМПЛЕКСНОГО ИССЛЕДОВАНИЯ

А.В.Быстров, Е.В.Злобин
(работа поддержана фондом "Культурная Инициатива")

Одна из целей изучения текстовых источников с использованием ПЭВМ - проведение автоматической атрибуции произведения, в авторстве которого у историков имеются сомнения. В разработке этих методов накоплен определенный опыт, выполнен даже историографический обзор имеющихся работ /1/.

Некоторое время назад одним из авторов был разработан для машин устаревшего класса, типа Искра-226, пакет прикладных программ для изучения текстовых исторических источников. Он позволял проводить ввод текста в машину, корректировку ошибок, автоматический подсчет более 40 статистических характеристик, морфологический анализ текста в диалоговом режиме. При этом была реализована методика Н.А.Морозова, основанная на подсчете частоты встречаемости служебных слов в предложении, дополненная рядом показателей на морфемном и лексическом уровнях (число слов разной длины в буквах и слогах, частота начальных букв слов и т.д.). Подсчитывалась также (после полуавтоматического кодирования) частота взаимной встречаемости грамматических форм в рамках простого предложения слева направо и на ее основе строился соответствующий граф и анализировалась его форма.

С использованием данного пакета был проведен анализ "Записок" декабриста И.И.Горбачевского, члена общества "Соединенных славян". Авторство "Записок" долгое время вызывало споры у ученых. Итоги подсчета статистически достоверно позволили сделать вывод о том, что автор "Записок", скорее всего, неизвестен.

Переработанный в программу для IBM-совместимых компьютеров данный пакет является одним из элементов автоматизированного рабочего места историка. Эта программа может работать с любым текстом, представленным в формате ASCII-файла. 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 конкретной версии не должно превышать 250. Возможен также поиск слов (словосочетаний), задаваемых исследователем. Количество их также может быть расширено. Главная цель создания программы - проведение атрибуции исторических источников.

Такая работа была проделана в рамках настоящего исследования при решении вопроса об авторстве предсмертного письма Б.В.Савинкова, известнейшего политического деятеля и талантливый писателя,

обстоятельства смерти которого на Лубянке в 20-х годах до сих пор не вполне ясны.

Вопрос об альтернативном авторстве предсмертного письма Б.В.Савинкова был, в частности, поднят А.И.Солженицыным в его опыте художественного исследования "Архипелаг ГУЛАГ" /2/. Автор приводит известное ему со слов М.П.Якубовича сообщение Я.Г.Блюмкина (видного левого эсера, "прославившегося" участием в убийстве немецкого посла, графа Мирбаха в Москве в 1918 г., сотрудника ВЧК - ГПУ) о том, что именно он, Блюмкин, написал знаменитое предсмертное письмо Б.В.Савинкова Ф.Э.Дзержинскому от 7 мая 1925 года /3/. Письмо явилось последним автографом Б.В.Савинкова, потому что после этого, при не до конца ясных обстоятельствах, он покончил жизнь самоубийством, выбросившись из окна.

Для правдоподобной имитации стиля и присущих Б.В.Савинкову эмоциональных особенностей письменной речи Я.Г.Блюмкин якобы вживался в образ, несколько раз по многу часов беседуя с Савинковым в его камере во внутренней тюрьме ГПУ. Предсмертное письмо было написано, по мнению Солженицына, чтобы представить убийство Савинкова как самоубийство, закономерно завершающее его полное идеологическое и политическое разоружение перед лицом Советской власти, глубокое раскаяние в содеянном, своеобразный первый удачный опыт "перековки" и публичного покаяния.

Возможность почерковедческой экспертизы имеющегося первоисточника представляется проблематичной, поскольку при выманивании Б.В.Савинкова из-за рубежа ГПУ использовало услуги почеркиста, который подделал почерк соратника Савинкова С.Э. Павловского.

Статистическая обработка текстов была проведена нашей программой. При этом исходный массив служебных слов был несколько видоизменен и дополнен с целью выявления особенностей, присущих текстам малого объема /4/.

Сравнительному исследованию подверглись 14 текстов Савинкова (которые можно условно подразделить на художественные и публицистические) /5/, 2 текста Блюмкина /6/ и собственно предсмертное письмо /7/.

По совокупности полученных статистических характеристик, присущих текстам Б.В.Савинкова, можно сделать вывод, что художественные произведения Б.В.Ропшина (его псевдоним), такие, как "То, чего не было", "Конь вороной" и пр., существенно отличаются от публицистических статей и произведений эпистолярного жанра (различного рода писем). Соединение этих двух массивов в один вызывает возрастание дисперсии признаков (как и ранее, в данном исследовании все параметры пересчитывались на 1000 словоформ) в 2-3 раза, что сразу же снижает возможность проверки гипотез с использованием критерия Стьюдента. Видимо, в данном случае жанровые отличия оказывают

существенное влияние и на частоту употребления различного рода слов, и на грамматический строй предложения.

В случае же дифференцированного подхода к источникам каждого жанра, выделения писем и показаний в отдельную группу, просчета статистических характеристик по текстам, в нее входящим, результаты анализа получаются совершенно иными. В этом случае меры изменчивости вариационных рядов (в данном конкретном примере речь идет о дисперсии) гораздо меньше. Это свидетельствует об устойчивости статистических характеристик текстов, принадлежащих одному и тому же автору. А так как последнее, предсмертное письмо Б.В.Савинкова (в авторстве которого и высказываются сомнения) относится именно к эпистолярному жанру, то вполне правомерно сравнение его характеристик именно с этим массивом.

При таком допущении оказывается, что по абсолютному большинству своих характеристик анализируемый текст ближе к текстам, принадлежащим Я.Блюмкину, чем к усредненным параметрам текстов Б.В.Савинкова. Однако близость эта перекрывается статистическими характеристиками рассеивания для данных вариационных рядов.

Количественная оценка степени достоверности каких-либо выводов в данном случае затрудняется ввиду крайне малого объема имеющегося в нашем распоряжении текстового материала. С одной стороны, сама записка мала по объему (всего 370 словоформ), с другой - единственные дошедшие до нас тексты Я.Блюмкина также невелики. Поэтому использование МЧД и просчет характеристик текста на ПЭВМ позволяют сделать лишь предположительное заключение по вопросу атрибуции этого источника.

Этот вывод хорошо согласуется и с данными глубокого содержательного анализа имеющихся источников обычными методами. Во-первых, посещал ли Блюмкин Савинкова в тюрьме? То есть, был ли он в это время в Москве, имел ли такую возможность и желание? В Москве он был, поскольку именно в связи с отставкой Троцкого /8/ с поста председателя Реввоенсовета (январь 1925) Блюмкин переходил из его (Троцкого) аппарата в ведомство ОГПУ.

В то же время, по свидетельству Г.Агабекова, "Блюмкин пользовался большим влиянием среди членов коллегии ОГПУ и стал "в своем роде знаменитостью" ... благодаря успешной деятельности в Монголии", т.е. по крайней мере к середине 1928 г, когда его командировали в Константинополь /9/. Его фамилия была даже включена в 6-й том БСЭ в апреле 1927 г. /10/

Блюмкина называли "признанным любимцем Дзержинского", - умершего в июле 1926-го, т.е. через год с небольшим после гибели Савинкова./11/ Видимо, Блюмкин был им и некоторое время прежде; то, что его взяли из аппарата опального Троцкого в ОГПУ, этому не противоречит. Ю.Фельштинский обратил внимание на то, что "дальность знакомства Дзержинского с Блюмкиным (на 6 июля 1918 года) сомнительна. Так,

нагрянув в отряд Попова в поисках убийцы Мирбаха, Дзержинский обнаружил удивительную осведомленность в подробности туалета Блюмкина, "вычислив" его по забытой на столе шапке"/12/.

То есть, во время заключения Савинкова в тюрьме Блюмкин пользовался достаточным влиянием в ОГПУ и административная возможность посетить заключенного у него была. Интерес Блюмкина к Савинкову был вполне возможен. Наш анализ предактного (перед убийством Мирбаха) письма позволяет выявить его возможного адресата - Савинкова /13/.

Есть и прямое свидетельство об авторитетности Савинкова для Блюмкина "поверх партийных барьеров", ставших еще выше в начале 1921 г., когда Блюмкин просил И.Г.Эренбурга узнать в Париже, как Борис Викторович смотрит на "уход с акта", т.е. должен ли террорист, убивший политического врага попытаться скрыться или предпочтительнее заплатить за убийство своей кровью /14/ .

Кроме того, ко времени заключения Савинкова в тюрьму у них с Блюмкиным появились еще общие с ним проблемы, посоветоваться по которым можно было только между собой: Савинков в августе 1924-го был арестован и согласился работать /15/ для примирения Советской России и российской демократической эмиграции, не выдав своих еще законспирированных товарищей по борьбе /16/ ; при этом все его друзья по эмиграции от него отвернулись; в апреле 1919 года разыскиваемый Блюмкин по собственной инициативе явился в киевскую губЧК, чтобы объяснить по поводу своего участия в теракте против Мирбаха и действий ЦК левых эсеров в событиях 6 июля 1918 года для спасения репутации партии в глазах Советской власти и примирения ее с большевиками /17/ ; при этом он отказался выдать своих товарищей по партии /18/, чему те не поверили, и на Блюмкина было совершено три покушения /19/.

Современники Якова Блюмкина характеризуют его как большого хвастуна, краснобая и любителя приврать, соглашаясь в этом с другими авторами /20/, хотя прямо и не относят это к анализу слов Блюмкина о написании письма Савинкова /21/, ими цитируемых. Однако, если их отношение к достоверности слов Блюмкина можно понять как скептическое, то в том, что такой рассказ был (такое вранье было), они не выражают ни малейшего сомнения; более того, их исследование можно понять именно как аргумент в подтверждение достоверности рассказа Якубовича, который "Блюмкину не верил, думал, что хвастается" /22/ . Отсюда естественно возникает вопрос: с чего бы Блюмкину пришло в голову хвалиться именно авторством (по его словам) предсмертного письма Савинкова? Что он знал о редактировании письма (что следует из современной версии, опубликованной в пробном номере журнала "СВ" (Служба безопасности)) /23/? К сожалению, из текста Солженицына не ясны обстоятельства, место и время (между маем 1925 и ноябрем 1929) беседы Блюмкина с Якубовичем.

Можно предположить, что Блюмкин хвастался, когда рассказ был актуальнее, т.е. ближе к первой дате.

Таким образом, анализ имеющихся сведений позволяет сделать вывод, что, по-видимому, Блюмкин посещал Савинкова во время заключения последнего во внутренней тюрьме ОГПУ. Скорее всего, это происходило в марте - начале апреля 1925 года.

И, скорее всего, Блюмкин, относившийся к Савинкову с большим пиететом, не мог написать за него предсмертного письма.

Этот вывод в целом совпадает с полученными статистическим путем данными. Математические методы и конкретно-исторический содержательный анализ источника должны взаимно обогащать и дополнять друг друга, способствуя получению новой, значимой информации.

ПРИМЕЧАНИЯ:

1. Чернов Е.А. Проблемы применения количественных методов при изучении авторских стилей в нарративных источниках по отечественной истории // Проблемы применения количественных методов анализа и классификации источников по отечественной истории. Днепропетровск, ДГУ, 1988, с.36-48.

2. См. Солженицын А.И. Архипелаг Гулаг // Новый мир, 1989, N 9, с.108-109.

3. См. Известия, 1925, 13 мая.

4. Список просчитываемых в текстах служебных слов составил 196 словоформ и был скорректирован по частицам, союзам, предлогам. См. Чешко Л.А. Русский язык. М., 1990, С.193, 195, 188; Русский язык. М., 1985, с.332.

5. Например, письменное показание Бориса Савинкова, данное им 21 августа 1924 года // Борис Савинков перед Военной Коллегией Верховного суда СССР. М., 1990, с.710 и др. источники.

6. Показания Я.Блюмкина от 17-19 апреля 1919 года, Киев // Красная книга ВЧК. М., 1990, т.1, с.295-305; Яков Блюмкин. Показания КГЧК 8 мая 1919 года, там же, с. 305-308.

7. Известия, 1925, 13 мая; Правда, 1925, 13 мая.

8. Овруцкий Л., Разгон А. Яков Блюмкин. Из жизни террориста. Горизонт, 1991, N.12, с.58.

9. Агабеков Г. ГПУ (Записки чекиста). Берлин, 1930, с.221.

10. Цит. по: Овруцкий Л., Разгон А. Указ. соч., N 11, с.49.

11. ЧК за работой. Берлин, 1931, с.295.

12. Овруцкий Л., Разгон А. Указ. соч., с.52.

13. Там же, с.57-58.

14. Эренбург И. Люди. Годы. Жизнь. М., 1990. Т.1. С.194.

15. См. его письма к друзьям по эмиграции: И.И.Фундаминскому (Б.Савинков. Посмертные статьи и письма. М., 1926. Библиотека

"Огонек", N.95, с.710), Д.С.Пасманнику (с.11-23), Д.В.Философову (Б.Савинков. Письмо Философову. Харьков, 1924, с.515), В.Л.Бурцеву (Савинков перед военной коллегией Верховного Суда. М., 1924, с. 263-266) и др.

16. Савинков Б. Тюремные дневники. 1 мая // Служба безопасности, М., Пробный номер, 1992, с.20.

17. Память. Исторический сборник. Вып.3. М.,1978. Париж; 1980, с.380.

18. Овруцкий Л., Разгон А. Указ. соч., С.53,64; Показания Я.Блюмкина // Красная книга ВЧК, с.302.

19. Овруцкий Л., Разгон А. Указ. соч., с.53.

20. Агабеков Г. Указ. соч. с.222; В.Шершеневич. Великолепный очевидец. Поэтические воспоминания 1910-1925 гг. // Мой век, мои друзья и подруги. Воспоминания Мариенгофа, Шершеневича, Грузинова. с.612-614; Показания М.Лациса // Красная книга ВЧК, с.264.

21. Солженицын А.И. Архипелаг ГУЛАГ // Новый мир, 1989, N.9, с.110.

22. Там же. с.109.

23. Служба безопасности. Пробный номер. М., 1992, с.19.

ОБ ИЗУЧЕНИИ СОЦИАЛЬНОЙ ДИФФЕРЕНЦИАЦИИ
ЗАКАВКАЗСКОГО КРЕСТЬЯНСТВА К КОНЦУ XIX в.
(по материалам бюджетного обследования)

И.Р.Рафи-заде

Важнейшей задачей историков-аграрников представляется исследование процесса социальной дифференциации крестьянства в различных регионах, выявление и характеристика различных социальных типов крестьянских хозяйств, общего уровня их развития. К сожалению, истинные масштабы развернувшегося процесса социальной дифференциации крестьянства известны весьма приблизительно. В историографии отсутствуют достаточно четкие и определенные представления о тех конкретных типах крестьянских хозяйств, которые сложились и функционировали к концу прошлого века в различных регионах, и в частности в закавказской деревне. Социальная структура закавказского крестьянства весьма скудно отражена в источниках и весьма слабо освещена в исторической литературе.

В настоящей статье нами предпринята попытка осветить данные вопросы, используя первичную массовую статистическую информации о крестьянском хозяйстве, на основе научной разработки этого материала в связи с созданием и обработкой базы данных.

Нами предпринята попытка научного анализа процесса социальной дифференциации закавказского крестьянства эпохи капитализма по массовым данным. Созданный с этой целью архив машиночитаемых данных (далее МЧД) хронологически охватывает 80-90-е гг. прошлого столетия. Основное ядро архива составил обширный статистический массив единственного дореволюционного обследования крестьянского хозяйства Закавказского края середины 1880-х гг., рассредоточенный в многочисленных публикациях и содержащий подробные и сопоставимые сведения о численном составе крестьянства, экономическом состоянии крестьянского хозяйства, податном обложении и недоимках, о структуре земельного фонда и др. /1/

Описание созданной базы МЧД, а точнее, коллекции баз данных по аграрной истории Закавказья XIX в., уже являлось предметом нескольких специальных публикаций /2/, а потому не имеет смысла возвращаться к данному вопросу здесь в полном объеме.

В территориальном плане база данных охватывает все Закавказье на поуездном уровне. Отметим лишь, что по каждому уезду, губернии, селению края был выработан единый, стабильный и, следовательно, сопоставимый круг показателей, описывающих экономическое, социальное, демографическое развитие закавказской деревни. При отборе показателей

мы руководствовались стремлением избежать потери информации, максимально мобилизовать информативные возможности изучаемых данных.

Самостоятельный блок банка данных образуют материалы крестьянских бюджетов Закавказья конца XIX в., впервые вводимые в научный оборот. Изучаемый массив включает подробнейшие сведения о наделе, посевах, пропорциях семейного и наемного труда, затратах на хозяйственные нужды и на личное потребление (в том числе на пищу, одежду, предметы домашнего обихода), а также данные о свободных денежных средствах, о потребляемой и товарной части продукции (в натуральном и денежном выражении) и др.

Рассматриваемая база первичных бюджетных данных о хозяйстве государственных крестьян охватывает все описанные в процессе обследования 1880-х гг. крестьянские хозяйства (всего 44 бюджета). Информация учитывалась во всем объеме, без ограничений и информативных потерь (всего содержание каждого бюджета включает 400 показателей). Общий объем базы данных составил свыше 17 тыс. первичных показателей по бюджетам. Вследствие богатства и достоверности данных бюджетные сведения позволяют углубить и конкретизировать анализ социально-экономического развития деревни, основанный на материалах сплошных подворных переписей хозяйств.

С обилием и разнообразием информации бюджетных обследований сопряжены объективные трудности работы с этим материалом. Вплоть до настоящего времени эффективность использования данной группы источников по аграрной истории России периода капитализма не соответствует ее огромному информативному потенциалу /3/. В качестве возможного подхода здесь представляется моделирование разнообразных социальных типов крестьянского хозяйства, образовавшихся в процессе социальной дифференциации закавказского крестьянства к концу прошлого века. Данная группа источников относительно немногочисленна. За весь период капитализма бюджетные обследования всего по России охватили около 12 тыс. крестьянских хозяйств /4/. Дореволюционные бюджетные обследования проанализированы в специальной литературе /5/.

В Закавказье бюджетная статистика значительного распространения не получила. В этом свете особо важное значение имеет корпус бюджетных описаний крестьянских хозяйств, созданный в процессе обследования государственной деревни края 1880-х гг. В настоящем исследовании впервые в отечественной историографии данные крестьянские бюджеты вводятся в научный оборот. База данных по крестьянским бюджетам создавалась совместно с кандидатом физ-мат. наук А.С. Гусейновой. На ее основе создавались обобщенные математические модели различных экономических групп крестьянских хозяйств.

Математическое моделирование в гуманитарных исследованиях получает все больший размах /6/, в том числе в исторических и историко-социологических работах /7/. Сферой традиционного применения такого

подхода можно считать анализ аграрного строя России XIX- начала XX в. При моделировании экономического функционирования различных категорий крестьянских хозяйств в качестве главной задачи выдвигается обоснованное решение проблемы выявления основных экономических и социальных типов крестьянских хозяйств, оформившихся в государственной деревне Закавказья к концу XIX в., для последующего углубленного сравнительного изучения выявленных экономических типов.

Представленность в изучаемом статистическом массиве всех основных групп (типов) крестьянских хозяйств заложена в самой идее проведения бюджетного обследования 1884-1886 гг. Собранный бюджетным обследованием сплошная информация в дальнейшем целенаправленно публиковалась для контрастных групп крестьянских хозяйств. Бюджетные обследования охватили подавляющую часть социальных типов крестьянских дворов /8/. По различным уездам дано бюджетное описание - до 400 и свыше подробных показателей - для "сильной" (зажиточной), "средней" и "слабой" или "бедной" семьи. В бюджетном обследовании представлены также различные оттенки как зажиточного, так и бедного и среднего экономических слоев, что хорошо видно из помещенного ниже статистического материала (табл. 1-4). Следовательно, по своему стремлению к выявлению разнообразных экономических типов крестьянских хозяйств (а не однородных хозяйств) обследование середины 1880-х гг. совпадает с целями и задачами настоящего исследования. Сказанное многократно усиливает информативность этого материала, ощутимо повышает содержательную репрезентативность бюджетов, несмотря на относительно небольшое их общее количество, что позволило моделировать на этом материале пять достаточно рельефных экономических групп крестьянских хозяйств, выделить и охарактеризовать основные производственно-экономические и социально-экономические типы крестьянских хозяйств (но не их количественное соотношение в общей массе крестьянства), раскрыть общие и особенные черты их экономической деятельности /9/.

В базе данных каждое хозяйство описывалось как замкнутая экономически функционирующая система. Экономическое функционирование крестьянского хозяйства предполагает последовательное прохождение строго определенных экономических процессов: посев, сбор урожая, личное потребление, уплата податей, приобретение предметов домашнего обихода, покупка сельхозинвентаря, ремонт построек и т.д.

Основная идея работы заключается в том, что замкнутая математическая модель любого производственно-экономического явления (в том числе крестьянского хозяйства) позволяет вычислить все внутренние величины (включая случайные пробелы источника), если все ее внешние величины правильно заданы /10/. На этом принципе основаны подсчет любых параметров для любой группы хозяйств, "перекачивание" информации из различных информационных блоков, "проигрывание" всех возможных

вариантов экономической (хозяйственной) группировки крестьянских дворов. В этом и состоит главное преимущество изложенного подхода в источниковедческом ракурсе, возможность реконструкции (вероятностной), восстановления многочисленных недостающих данных, пробелов информации.

Чтобы правильно ориентироваться во всем многообразии бюджетных данных и максимально объективизировать, оградить от исследовательского произвола процесс вычленения экономических групп крестьянства, была подготовлена специальная программа упорядочения показателей, позволяющая расположить хозяйства по убывающей (от самого большого к самому меньшему) или по возрастающей (последовательно от меньшего к большему) для любого желаемого экономического показателя.

Анализ полученных результатов представляет большой интерес при оценке сущности экономических различий между типами хозяйств, для уточнения общих представлений об экономическом состоянии и функционировании крестьянского хозяйства.

Рассмотрим степень нарушения экономической однородности крестьянства, размытости социальной структуры государственной деревни на крайних (наименьших и наибольших) показателях общей стоимости имущества крестьянской семьи и его составных (стоимость скота, дома и др.), годовых доходов, расходов хозяйств и др. Общая стоимость имущества крестьянских семей колебалась в колоссальной амплитуде: от 20 руб. в наиболее малоимущих семьях и до 13 тыс. - в наиболее зажиточных. Не только по этому комплексному показателю, но и по всем составным компонентам имущественной состоятельности наметились полярные экономические типы. Стоимость скота (в руб.) колебалась от 0 руб. 00 коп. в нескольких безлошадных хозяйствах и 30 руб. - в беднейших до 8500 руб.- в зажиточных хозяйствах. Стоимость дома - от 10 руб. в двух фактически батрацких хозяйствах и до 2-5 тыс. руб. - в зажиточной группе.

Что касается производственной базы (стоимость сельхозинвентаря), то здесь различия хотя и весьма ощутимы, но не столь контрастны. Если в двух низших группах этот показатель равен нулю, то и в высших группах стоимость сельхозинвентаря нигде не поднимается выше 80 руб., что свидетельствует о недостаточном внимании к развитию производственно-технической базы в крестьянском хозяйстве в целом.

Уровень доходов крестьянской семьи колеблется весьма существенно: от 40 руб. (следующий показатель 120 руб.) в беднейших хозяйствах и до 8200 руб. (предыдущий показатель 3700 руб.) - в зажиточных. Уровень доходов на душу в беднейшей семье составлял всего 8 руб. (следующий показатель 17 руб. 30 коп.) и достигал свыше 300 руб. на каждого члена семьи в зажиточных хозяйствах.

Более высокий уровень доходов диктовал и более высокий уровень расходов, их особую структуру, в частности более высокие затраты на личное потребление (пища, одежда) и др. Всего на душу крестьянского

населения сумма годового расхода колебалась от 9 руб. (следующий показатель 19 руб. 30 коп.) в беднейших семьях и до более чем 200 руб. расходов на душу в зажиточных семьях.

Даже если отбросить эти наиболее рельефные полярные данные и рассматривать данные в среднем, умеренном интервале ("центр тяжести"), то по всем важнейшим экономическим показателям различия между хозяйствами остаются очень значительными. Общая сумма доходов (без учета крайних групп хозяйства) изменяется от 100-120 руб. до 1000 и выше. Общая стоимость имущества (также без крайних данных) возрастает от 100 руб. и до 1-3 тыс., стоимость скота - от 30 руб. и до 1-2 тыс. и т. д. Все это однозначно свидетельствует о достаточно определившихся к концу века качественно различных сформировавшихся экономических слоях государственных крестьян, о необратимой экономической неоднородности, которую достаточно ярко отразили бюджетные данные. Правомерно, следовательно, выделение и изучение на этом материале социально-экономических групп крестьянских хозяйств.

Задача обоснованного выделения экономических групп крестьянских хозяйств не так проста, как это кажется на первый взгляд. Варианты здесь, по сути, не ограничены, так как имеется реальная возможность группировки хозяйств по 400 различным признакам. Для решения подобных задач используется метод аналитических группировок, когда значения одного признака (группировочного) разбиваются некоторым образом на интервалы (например, по величине посева: до 2-х дес., от 2-х до 5-ти дес., от 5-ти до 10-ти дес. и т.д. или по числу голов скота в хозяйстве: 1, 1-2, 2-3, 3-4 и т.д.). Затем для образования групп (в результате произведенной таким образом группировки) по каждому интервалу подсчитывается среднее значение другого признака (или целого ряда интересующих признаков или всех вообще известных признаков).

В связи с этим возникает сразу два принципиальных вопроса. Первый момент: выделение основного (или основных) группировочных признаков, способных отразить качественные различия между группами крестьянских хозяйств. С другой стороны - это определение интервалов, установление той грани, которая отделяет экономические группы хозяйств. Общепринятых критериев подобного рода (за исключением лишь обеспеченности землей и скотом) в историографии пока не выработано.

Здесь возможен и предпочтителен многоплановый, гибкий подход. При разбивке хозяйств на экономические группы (слои) мы не исходили из жестких, формальных критериев, а придерживались гибкого многопланового классификационного принципа. Формирование экономических групп было организовано в диалоговом режиме и представляло собой процесс "проигрывания" возможных вариантов группировки по различным группировочным признакам. Процедура группировки обеспечивалась подсистемой формирования запросов, осуществляющей поиск в базе данных по нескольким признакам (с многоступенчатой кодировкой и подсистемой

выдачи отобранной таким образом информации соответствующих бюджетов). В качестве основного группировочного признака выступало количество обрабатываемой земли (размеры посева) в хозяйстве. Остальные группировочные признаки носили вспомогательный характер.

Придерживаясь традиционной группировки, изучались следующие группы хозяйств: I - до 2 дес. обрабатываемой земли в хозяйстве; II - от 2 до 5 дес., т.е. две беднейшие группы; далее две средние группы: III - от 5 до 10 и IV - от 10 до 15 дес. Наконец, собственно зажиточный слой, V - обрабатывающий свыше 15 дес. земли. При этом учитывалось, что к мелкобуржуазным относятся хозяйства, обрабатывающие до 20 дес./11 / земли, а потому в ряде случаев хозяйства с посевом свыше 20 дес. рассматривались отдельно.

В исследовательской практике невозможно учесть весь контекст социально-экономического развития, не прибегая к системному анализу. Все гораздо сложнее и ближе к реальной действительности, когда социально-экономическое развитие (в том числе крестьянского хозяйства) изучается с учетом множества составляющих его элементов в их взаимодействии, что и подвело историков к системному анализу.

Используя этот подход для получения характерных хозяйств того или иного типа, кроме размера посева учитывались (на основе программы, построенной по принципу интервалов): общая стоимость имущества, жилых и хозяйственных построек, норма личного потребления, запас одежды, общая сумма доходов и расходов, затраты на сельхозинвентарь, расход на наемный труд и другие производственные нужды. Выдача всей этой информации по заданным параметрам, а также сортировка бюджетов по любому интересующему нас показателю, по сути, заложены в самой структуре базы данных, описывающей функционирование каждого крестьянского хозяйства с учетом мельчайших подробностей. Число заданных параметров можно увеличивать до 7(10), т.е., формируя облик хозяйства по всем важным параметрам, запрашивать нужные бюджеты, образующие, таким образом, данный экономический слой, прямо из памяти ЭВМ.

Примечательно, что при введении дополнительных (кроме основного признака - размера посева) характеристик, группировочных признаков и при повторной разбивке бюджетов по различным классам принадлежность хозяйства к "своей" группе, как правило, не изменялась. Менялось лишь расположение хозяйств внутри группы в зависимости от критериев сортировки бюджетов. Например, при разбивке хозяйств по стоимости жилых построек хозяйства "а", "b" и "с" располагаются в следующем порядке: "а" (N 1), "b" (N 2) и "с" (N 3). Те же хозяйства по запасу одежды (в денежном выражении) располагаются так: "а" (N 3), "b" (N 2) и "с" (N 1), а по стоимости скота: "а" (N 3), "b" (N 1) и "с" (N 3) и т.д. Разумеется, приведенный пример носит чисто условный характер. Состав же экономических групп в целом оставался стабильным.

Наряду с моделированием по вертикальному срезу - по экономической мощи - традиционные (бедные, средние, зажиточные) экономические типы можно формировать описанным путем "по горизонтали": земледельческий тип (с доходами от земледелия не менее 70% всех доходов), скотоводческий, садово-огороднический тип, а также изучать хозяйства с заработками не менее 30-50% "на стороне", вне хозяйства. Каждый сформированный таким образом экономический тип является как бы усредненным представителем своей группы, усредненным (по всем 400 известным параметрам) хозяйством данного экономического типа. Математическая модель описывает характерный для каждого социального слоя процесс производства, личного потребления, нормы товарности, прибыли, структуры доходов и расходов, стоимости хозяйственных и жилых построек, сельхозинвентаря и т.д.

Полученные обобщенные модели рассчитаны на интегральную характеристику различных социальных типов крестьянских хозяйств, глубоко и тонко отражают специфику их функционирования. Обобщенная модель учитывает характерные для каждой данной группы (типа) хозяйств соотношения экономических элементов. Набор хозяйств, включаемых в группы, можно многократно изменять по мере необходимости.

Полученные модели основных социально-экономических типов хозяйства государственных крестьян созданы на основе конкретного реального фактического материала, а потому распространение полученных результатов на государственную деревню в целом может носить до известной степени вероятностный характер. Проведенное исследование бюджетного материала не претендует на чрезмерные обобщения и расширительное толкование получаемых результатов и выводов.

В результате проведенного исследования были выявлены следующие группы крестьянских хозяйств (см.табл. I). По всем основным экономическим характеристикам полученные группы крестьянских хозяйств обнаруживают существенные отличия. Все рассматриваемые характеристики крестьянского хозяйства последовательно возрастают от низшей группы к высшей.

В высшей - V группе хозяйств общая стоимость имущества (1929 руб.) почти в 13 раз выше, чем в беднейшей - I группе (152 руб.). Стоимость как хозяйственных, так и жилых построек у зажиточного слоя почти в 11-12 раз превышает данный показатель у низшей части государственного крестьянства.

По размерам хозяйства и по структуре землепользования высшая - V группа коренным образом отличается как от двух беднейших, так и от двух средних групп. Если в беднейших хозяйствах размер посева в среднем достигал 0,5 десятины (что безусловно ниже прожиточного минимума), то площадь посева у наиболее обеспеченного V слоя достигает 32,96 дес. Это не только существенно выше посева обеих бедняцких

групп, но и в 3-6 раз превышает обрабатываемую земельную площадь обеих средних групп (III и IV).

Таблица I
Экономическая состоятельность различных
групп крестьянских хозяйств

Показатели	Группы хозяйств				
	I	II	III	IV	V
1. Размеры хозяйства (в дес.)	1.39	7.74	11.91	19.49	46.17
2. Площадь обрабатываемой земли (в дес.)	0.57	3.42	6.92	12.39	32.96
3. Арендваемой земли (в дес.)	0.05	0.50	0.78	1.00	9.67
4. Обрабатываемой земли на душу (дес.)	0.12	0.52	1.02	1.26	2.34
5. Общая стоимость имуще- щества (в руб.) 1929.2	152.04	438.38	658.5		2369.3
6. Стоимость дома (жилые постройки)	70.00	209.25	330.66	903.12	847.5
7. Стоимость хозяйствен- ных построек	21.14	51.11	63.18	152.50	290.8
8. Стоимость домашнего скота и птицы	176.09	172.24	261.28	644.52	315.3

Именно в этом, наиболее обеспеченном слое размеры арендованной земли достигают 10 дес. и выше, что составляет приблизительно 1/3

общего землепользования этих хозяйств. В остальных же четырех группах средний уровень аренды колеблется (с последовательным возрастанием по группам) от 0,05 до 1,00 дес. всего. Наблюдается падение основной роли надела именно в зажиточных хозяйствах, которые переносят центр тяжести с надела на вненадельную землю.

Такое же пропорциональное соотношение сохраняется между группами, если брать важнейшие экономические показатели – стоимость имущества, сумма расходов и доходов (годовой оборот) и т.д. не только в целом в хозяйстве, но и в расчете на каждого члена крестьянской семьи.

По уровню, по глубине различий кроме межгрупповых имеются еще и внутригрупповые колебания признаков. Если в одной группе объединены хозяйства, обрабатывающие от 5 до 10 дес. земли, то, само собой разумеется, хозяйство с посевом в 5,5 дес. существенно отличается от хозяйства с посевом в 9,5 дес. по всем экономическим характеристикам. В связи с этим в процессе исследования наряду с понятием "слой" использовалось также понятие "подслой", градации между которыми были, разумеется, более размытыми, менее рельефными. В этом смысле интервалы между группами в значительной мере носят условный характер. При помощи осуществленной группировки преследуется цель выявить именно существенно отличные по хозяйственной состоятельности группы крестьянских дворов. Такой подход позволяет изучать специфику и сходство производственно-экономического функционирования, внутреннего строя различных типов хозяйств в зависимости от глубины экономических различий; установить, насколько выявленные различия между отдельными экономическими группами крестьянских хозяйств носили качественный характер. Именно этой задаче отвечает осуществленная экономическая группировка. Полученные же 5 групп крестьянских хозяйств представляют собой, как мы видим, достаточно рельефно выраженные хозяйственно-экономические типы, что и требуется в данном случае.

Особенно отчетливо ощущается экономическое неравенство между различными типами крестьянских хозяйств при анализе доходов и расходов (валовых и дифференцированных) у различных слоев крестьянства (см. табл. II). Зная уровень доходов и расходов каждого социального слоя государственной деревни, можно дать качественную оценку степени "размытости" социальной инфраструктуры, глубины различий между отдельными группами крестьянских хозяйств.

С этой целью вводится термин "норма накопления", "запас слоя", что очень важно в плане оценки возможностей обогащения и разорения, потенциальных перспектив развития той или иной группы крестьянских хозяйств. Речь идет о выявлении основных тенденций социального развития. Валовые доходы, а также расходы крестьян систематически возрастают от низшей группы к высшей, отличаясь в I и V группах более чем в 15 раз. В том числе, доходы от земледелия превосходили здесь в 42 раза (а по отношению к предыдущей IV группе – в 17 раз), доходы от

скотоводства - в 14 раз. При этом отмеченный контраст сохраняется и в расчете на каждого человека. Каждый член семьи в хозяйствах I группы имел в среднем 33 руб. годового дохода, а расходовал около 35 руб., и, следовательно, бюджет его был дефицитным.

II группа в общем дефицитной не являлась, жизненный уровень здесь был несколько выше. Между тем каждый представитель данного слоя жил, что называется, едва сводя концы с концами, имея в среднем около 60 руб. годового дохода, расходовал ежегодно около 54 руб. и, следовательно, имел 6 руб. чистого дохода в год. Это означает, что лишь небольшая часть хозяйств данной группы имела шансы удержаться на грани разорения. В целом же подавляющая масса хозяйств данного типа имела отчетливую нисходящую перспективу развития, эволюционировала по направлению к I группе со всеми ее вышеприведенными характеристиками.

Таблица II
Доходы и расходы различных групп
крестьянских хозяйств (в руб.)

Показатели	Группы хозяйств				
	I	II	III	IV	V
1. Валовой доход семьи	210.10	381.83	526.87	1060.4	
3206.08					
2. Доход от земледелия	46.91		251.64	423.13	689.79
1697.82					
3. Доход от					
скотоводства	87.05	50.88	68.46	166.34	
1272.18					
4. Валовой доход					
на душу	33.73	58.79	74.12	100.90	
245.09					
5. Валовые расходы					

семьи	213.10	365.80	456.29	854.23	2402.89
6. Валовой расход					
на душу	34.94	53.78	63.64	80.70	199.73
7. Чистый доход семьи	-3.14	19.04	75.56	217.05	
837.09					
8. Чистый доход на душу	-0.46	6.85	10.55	20.19	45.35

Следующие, III и IV группы в среднем имели доход на одного человека в год 74 руб. и 100 руб., расходовали соответственно 63 руб. и 81 руб. Чистый доход, или норма накопления, данного слоя 10,5 и 20,1 руб. ежегодно (у IV группы в два раза выше, чем у III). Хотя обе эти группы представляют среднее крестьянство государственной деревни, как видим, между ними имелись существенные экономические отличия, в том смысле, что положение IV слоя было гораздо более устойчивым, нежели III. Большая рентабельность и прибыльность хозяйств этого типа (IV группы) создавали реальную перспективу перехода определенной части хозяйств в следующую, высшую группу. Норма чистой прибыли здесь весьма высока (в среднем ежегодно до 217 руб. накопления).

И еще одно наблюдение. Хотя все группы достаточно рельефны, разрыв между ними нарастает с переходом к каждой следующей группе. Наиболее резкий разрыв - по всем экономическим показателям - наблюдается между IV и V группами хозяйств.

Хозяйства высшего типа (V группа) располагали значительными денежными средствами, характеризовались большим годовым оборотом капитала при годовых доходах в среднем 3206 руб., в том числе от земледелия - 1697 руб. и от скотоводства - 1272 руб., с валовым расходом 2402 руб., с нормой чистой прибыли в хозяйстве 837 руб. и нормой накопления на каждого человека до 45 руб. ежегодно.

В целом поляризация социальной структуры была весьма глубокой. Различия хозяйств высшего и низшего слоя государственных крестьян практически по всем экономическим характеристикам были контрастными, с перерастанием в качественные социальные противоположности.

Уровень благосостояния и размеры крестьянских хозяйств, как нетрудно заметить, не зависят от числа трудоспособных членов крестьянской семьи, а также от общего числа домоладцев. Все аспекты экономической состоятельности зависели собственно от размеров и мощи крестьянского хозяйства, что является важным проявлением развития

буржуазных отношений в крестьянском хозяйстве. Более резкий скачок всех экономических показателей при переходе от IV группы хозяйств к V отражает те истинные качественные отличия, которые объективно существовали между производящим мелкотоварным, мелкобуржуазным хозяйством и собственно буржуазно-капиталистическим хозяйством. Мелкобуржуазное хозяйство характеризовалось, как известно, некоторыми отличительными чертами. Попытаемся обозначить сферу мелкотоварного хозяйственного уклада в рассматриваемых группах и конкретизировать, таким образом содержание мелкотоварного производства в крестьянском хозяйстве региона.

Прежде всего "мелкое товарное производство захватывает... несколько более обеспеченные слои крестьянской бедноты (8-15 дес. земли на двор и 1 лошадь) и средние дворы, ведущие в основном некапиталистическое хозяйство (15-20 дес. земли и 2-3 лошади на двор)" /12/. Поскольку мы здесь оперируем не общими размерами хозяйства, а только обрабатываемой землей, грани несколько смещаются. Это относительно обеспеченный слой - от 5 до 15 дес. землепользования, т.е. обе средние группы (III и IV группы). Вместе взятые, эти группы составляли наиболее многочисленную часть государственных крестьян (в Закавказье - 44,8%, в том числе III группа - 37,8% и IV - 19,96%; в Азербайджане - 48,7%, в том числе III группа - 30,4% и IV - 18,3%).

В то же время и по высшей границе трудно провести резкую черту (в действительности, без специального моделирования "чистых" типов) между высшим слоем мелкобуржуазных товарных хозяйств и собственно буржуазно-капиталистическими хозяйствами. В целом это, как правило, товаропроизводящие хозяйства с землепользованием не выше 20 дес. посева на двор /13/.

Наряду с этим чисто количественным показателем (площадь землепользования) мелкотоварное хозяйство характеризуется важными производственно-экономическими особенностями. Оно, в отличие от собственно буржуазного, капиталистически-предпринимательского, основано почти исключительно на собственном труде (не говоря уже о теснейшем переплетении личного труда с наемным также и в большинстве зажиточных хозяйств), а также отмечено крайней неустойчивостью и затрудненностью технического развития /14/.

Итак, в понятие мелкотоварного производства укладывается все пореформенное государственное крестьянство, исключая пролетаризированный слой, и собственно крестьянская буржуазия. Таким образом, при всей сложности социальной инфраструктуры государственной деревни края в основном крестьянское хозяйство представляло собой мелкобуржуазное товарное производство, которое "нечетко отделяется от полупролетаризированных хозяйств и оно же рядом незаметных переходов сливается с буржуазией" /15/.

Главным источником доходов и сферой деятельности государственных крестьян края было экстенсивное земледелие - полеводство - хлебопашество. Наиболее часто встречаются хозяйства с уровнем доходов от полеводства свыше 60-70% от общей суммы валовых доходов. Заработки "вне хозяйства" преимущественно были развиты в наименее и наиболее обеспеченных группах хозяйств. Если для первых это в основном поденная работа, а также отхожие заработки, то в состоятельном слое это торгово-промышленная деятельность (мельница, магазин, винный заводик и др.) либо государственное жалованье (сельский староста и др.).

Имеющиеся бюджетные данные позволяют, следовательно, проследить предельное развитие крайних социальных типов. Экономическое положение "чистой буржуазии и чистого пролетариата" отражают всего несколько изучаемых бюджетов - как оно и было (в пропорциональном отношении) в реальной действительности в рассматриваемый период. Чтобы представить размах хозяйства, имеющего до 30 десятин посева, заметим, что расходы по обработке 1 десятины пшеничного посева составляют по одним расчетам 29 руб. 95 коп. (т.72, ч.1, с.197), а по другим - 21 руб. (т.IV, ч.1, с.87) и где "годовой доход семьи простирается примерно до 3665 руб. 90 коп., расход до 3484 руб. 87 коп." (т.II, ч.2, с. 229). Бюджеты же "беднейших семейств" составляют разительный контраст и дают представление о тех, кто вынужден был, забрасывая свое хозяйство, поддерживать существование продажей своих рабочих рук.

Именно в этих двух противоположных направлениях и устремлялась перспектива развития остальной, срединной части хозяйств в процессе дальнейшего расслоения деревни.

Рассмотрим обеспеченность различных групп крестьянских хозяйств производственными фондами, а также пропорциональное соотношение производственных затрат и нормы прибыли на эти затраты в различных группах (см. табл. III).

Таблица III
Производственные фонды и затраты в
различных группах крестьянских хозяйств

Показатели	Группы хозяйств				
	I	II	III	IV	V

I. Что касается плуга (качественный показатель земледелия), то в первых двух группах плуг отсутствует во всех хозяйствах, в III группе применяется эпизодически, и только в двух высших группах - относительно регулярно.

Важный экономический показатель - использование наемного труда в крестьянском хозяйстве. Известны различные формы найма. Годовой (круглогодичный) найм в основном практиковался в наиболее зажиточных, кулацких хозяйствах. Гораздо шире был распространен сезонный, срочный найм, который был весьма развит не только в высшей группе крестьянских хозяйств, но и в подавляющем большинстве средних.

В бюджетах состоятельных семей обязательно имеет место указание на наличие нескольких наемных рабочих (как, например, "семья имеет надел из 7 десятин пахотной земли ... и арендует у татар до 12 десятин пахоты ... нанимает четырех постоянных рабочих из татар", или "нанимает двух работников для услужения дома и 14 пастухов" (т.VI, с.79-80; т.II, ч.I, с.358; т.2, ч.2, с.160; т.2, ч.2, с. 222). Обратимся к данным таблицы III. Все формы найма здесь учтены и даны в денежном выражении (в руб.). В III группе можно говорить о применении найма эпизодически. Уже в IV группе привлечение наемной рабочей силы представляет собой достаточно распространенное, обычное явление (80-90 руб. затрат в год). Крестьяне же V группы ведут свое многократно расширенное хозяйство на основе систематического использования наемного труда (до 250 руб. затрат на наемных рабочих в среднем в год).

Если соотносить объем податей с общими экономическими возможностями каждого хозяйства, то выясняется, что в высших группах подати составляют относительно небольшую долю (%) к общему годовому обороту и валовым доходам хозяйства. У крестьянской же бедноты уплата податей поглощала почти все экономические ресурсы хозяйства (за исключением личного потребления).

Анализ структуры податей показывает, что натуральная часть составляла ничтожную долю всех податей. Что же касается денежной подати, то она присутствует, во-первых, во всех бюджетах и колеблется от 4 руб. до 152 руб. Денежная рента выступает в государственной деревне не только как инструмент налогообложения, но и выполняет роль регулятора товарного производства. Значительная часть вырученных средств погашалась рентой, и значительная часть товарной массы выносилась на товарный рынок ради получения средств на уплату государственных податей. В результате крестьянские хозяйства всех групп были тесно и непосредственно связаны с рынком.

В то же время для низшей группы хозяйств (а в известной мере и для мелкотоварного производства вообще) вынос своей продукции на рынок носит вынужденный характер, и при этом подавляющая масса производимых в хозяйстве продуктов предназначена для личного потребления. Если

проследить судьбу той части сельскохозяйственной продукции, которая шла из малообеспеченных крестьянских хозяйств на продажу, то, во-первых, это была относительно небольшая часть; во-вторых, относительно небольшая часть вырученных денег шла на личное потребление. Основная же часть выбрасываемой из низшего слоя крестьянских хозяйств массы товара продавалась ради выручки средств для уплаты податей. В этом смысле централизованная рента выступала как регулятор рынка. Строго говоря, о товаропроизводящем мелкобуржуазном хозяйстве можно говорить начиная с III группы.

Как было показано выше, общая сумма производственных затрат систематически возрастает от низших групп к высшим, пропорционально размерам хозяйства. В результате размеры капиталовложений, материально-техническая обеспеченность труда в зажиточных хозяйствах оказывается качественно иной.

И если состоятельные слои деревни обеспечивали общественно необходимую прибыль (с каждого рубля затрат, с обрабатываемой десятины земли) за счет уровня капиталовложений, производственных затрат, привлечения посторонней наемной рабочей силы, то деревенская беднота, напротив, - за счет колоссального напряжения и перенапряжения своего собственного труда. По подсчетам И.Д.Ковальченко, беднейшему крестьянству для получения тех же доходов приходилось вкладывать примерно вдвое больше труда, чем крестьянам зажиточным /16/. Сказанным определялись разнонаправленная перспектива развития этих противоположных социальных типов крестьянских хозяйств.

Экономические различия между выделенными типами крестьянских хозяйств находили выражение в таких показателях, как общая доходность крестьянского хозяйства, доходность каждого члена семьи, доходность на всю площадь обрабатываемой земли, что и отражает объективное экономическое положение той или иной группы крестьянства (см. табл. II).

Глубокое проникновение рыночных отношений в ткань крестьянского хозяйства не могло не отразиться на производственно-экономическом строе этого хозяйства. Важным показателем глубины связи товаропроизводителей с рынком и подчиненности их хозяйства господствовавшим на нем закономерностям, а следовательно, и показателем степени втянутости этих хозяйств в товарно-капиталистические отношения, является возможность получения одинаковой прибыли на равные затраты производственного капитала /17/.

Этот показатель действительно сближается для всех крестьянских хозяйств - при столь контрастных всех прочих экономических показателях, что является важным подтверждением буржуазной природы развития хозяйства государственных крестьян в рассматриваемый период. Сказанное, разумеется, не означает, что экономические возможности всех крестьян были равнозначны - отнюдь нет. Напротив, и уровень товарности, и уровень доходности, а вместе с ним и все показатели

благосостояния весьма существенно отличались в зависимости от размеров хозяйства и резко контрастировали у высшей и у низшей группы (см.табл.II).

Как нетрудно заметить, с возрастанием размеров хозяйства возрастают и его доходы (в IV группе почти в три раза выше, чем в I). Доля производственных затрат и их сумма в высших группах была, естественно, наиболее высока, с чем и связано многократное возрастание здесь валового дохода. Анализ производственных личных затрат государственных крестьян, таким образом, отчетливо выявил базисные различия в структуре капитала, в производственно-экономическом строе разных типов крестьянских хозяйств.

И наконец рассмотрим личные затраты и норму личного потребления в различных группах крестьянских хозяйств (см.табл.IV). Крестьяне высшей группы (V) потребляли продуктов (и в целом, и на душу) почти в три раза больше, чем крестьяне низшей (I) группы.

Таблица IV

Личное потребление в различных группах
крестьянских хозяйств (в руб.)

Показатели	Группы хозяйств				
	I	II	III	IV	V
1.Личное потребление на					
душу	18.72	35.53	38.27	40.59	63.92
2.Потребление пищи на					
душу	14.45	27.86	29.84	29.56	45.68
3.Затраты на одежду на					
душу	4.27	7.67	8.43	11.03	17.24
4.Расходы на питание в					

хозяйстве		92.83		191.87		205.24		320.43	
606.88									

В то же время по соотношению с общим валовым доходом расходы на личное потребление (%) у малообеспеченных крестьян были гораздо выше, чем у состоятельных.

До 90 % всех доходов, получаемых колоссальным напряжением собственного труда (I и II группы), приходилось тратить на личное потребление (сюда в основном входит натуральная часть, все то, что не выносилось на рынок, а потреблялось в самом хозяйстве). Здесь объективно не оставалось экономических ресурсов на развитие производственно-технической базы хозяйства, на его расширение и др. Все сказанное обнаруживает значительно худшее положение и отсутствие реальных возможностей накоплений, а следовательно, и перспектив развития для социальных низов деревни, деревенской бедноты.

* * *

Проведенное исследование впервые позволило наполнить конкретным содержанием традиционно известные "зажиточную", "бедную" и "среднюю" группы (экономические слои) крестьянства, количественно и качественно оценить и соизмерить глубину различий между ними.

Предложенный путь анализа позволяет поднимать новые пласты информации для комплексной характеристики различных социальных типов крестьянских хозяйств и общего уровня их социально-экономического развития. Проведенное исследование создает основу для целостного системного изучения разложения закавказского крестьянства в эпоху капитализма в общерегиональном масштабе.

ПРИМЕЧАНИЯ

1. Материалы для изучения экономического быта государственных крестьян Закавказского края. Тифлис, 1885-1887. тт.1-7, Статистические приложения; Сборник данных о землевладении и способах хозяйства в пяти губерниях Закавказского края, Тифлис, 1899. Приложение к записке уполномоченного министра государственных имуществ на Кавказе Я.С.Медведева, представленной в комиссию по составлению проекта законоположений о поземельном устройстве государственных поселян Закавказского края // Антелава И.Г. Государственные крестьяне Грузии в XIX веке, т.2, Тбилиси, 1962; Свод статистических данных о населении Закавказского края, извлеченных из посемейных списков 1886 г. Тифлис, 1893; Свод статистических данных, извлеченных из посемейных списков населения Кавказа. Тифлис, 1888; Сборник статистических сведений по

Закавказскому краю. Тифлис, 1902; Первая всеобщая перепись населения Российской империи 1897; Общий свод по империи результатов разработки данных первой переписи населения, проведенной 28 января 1897 г. Спб., 1905; Кавказ в сельскохозяйственном отношении в 1899 г. Тифлис, 1902; Сборник статистических сведений по Закавказскому краю. Тифлис, 1902; Краткие сведения о состоянии на Кавказе сельского хозяйства. Тифлис, 1900; Сборник сведений по плодоводству в Закавказском крае, вып. I. Тифлис, 1895; Закавказская железная дорога. Статистика перевозок грузов, отправляемых по Закавказским железным дорогам в 1884-1903 гг. Тифлис, 1885-1904 и др.

2. Подробнее об этом см.: Рафи-заде И.Р. О создании машиночитаемого банка данных по аграрной истории Закавказья эпохи капитализма // История ремесла и промышленности Азербайджана в период позднего феодализма и генезиса капитализма. Баку, 1990; Рафи-заде И.Р., Гусейнова А.С. Об изучении материалов бюджетного обследования крестьянских хозяйств Закавказья конца XIX в. в режиме базы данных // Информационный бюллетень комиссии по применению ЭВМ в исторических исследованиях, IX выпуск. М., 1993; Irada R. Rafizade. About the Database Collection of Machine-Readable Statistical Information on the Socio Economic History of Azerbaijan // VIII International ANS Conference, Graz, Austria, August, 24-27, 1993, Abstracts.

3. Ковальченко И.Д., Селунская Н.Б., Моисеенко Т.Л. Социально-экономический строй крестьянского хозяйства России в эпоху капитализма. М., 1988, с.39.

4. Ковальченко И.Д., Селунская Н.Б., Моисеенко Т.Л. Указ. соч., с.89.

5. См.: Чайнов А., Студенский Г. История бюджетных исследований. М., 1922; Чайнов А.В. Бюджетные исследования. История и методы. И., 1929; Корневская Н.Н. Бюджетные обследования крестьянских хозяйств в дореволюционной России. М., 1954 и др.

6. Штофф В.А. Роль модели в познании. Л., 1963; Он же. Моделирование и философия. М.-Л., 1966; Глинский Б.А. Моделирование как метод научного исследования. М., 1965; Бирюков Б.В., Геллер Е.С. Кибернетика в гуманитарных науках. М., 1973; Уемов А.И. Логические основы метода моделирования. М., 1971; Мамедов Н.М. Моделирование как синтез знаний. Баку, 1978; Яглом И.М. Математические структуры и математическое моделирование. М., 1980; Батароев К.Б. Аналогии и модели в познании. М., 1981; Неуймин Я.Г. Модели в науке и технике. История, теория, практика. Л., 1984.

7. Филатов Д.А. Методологические вопросы моделирования объектов прошлого, настоящего и будущего в социальной науке. Саратов, 1977; Математика в социологии. Моделирование и обработка информации. М., 1977; Цыба Т.В. Математико-статистические методы в социологическом исследовании. М., 1981; Молчанов В.И. Системный анализ социологической

информации. М., 1981; Завьялова М.П. О моделировании в историческом исследовании. Автореф. дисс. ... канд. ист. наук. Томск, 1970; Она же. Метод ретроспекции и моделирование в историческом исследовании // Вопросы методологии истории и историографии. Вып.3, Томск. 1974; Товма Т.Д. Роль моделирования в познании социально-исторических явлений // Методологические вопросы науки. Вып.5. Саратов, 1977; Ковальченко И.Д. О моделировании исторических явлений и процессов // Вопросы истории, 1978, N 8; Устинов В.А., Кузищин В.И., Павловский Ю.Н., Гусейнова А.С. Опыт имитационного моделирования историко-социального процесса // Вопросы истории, 1976, N 11; Гусейнова А.С., Павловский Ю.Н., Устинов В.А. Опыт имитационного моделирования. М., 1984 и др.

8. Любопытно, что подсчет удельного веса (%) групп крестьянских хозяйств по обрабатываемой земле по изучаемым бюджетам совпал в процентном отношении не только с общими их пропорциями в крае, но и с полученными по другому региону (Европейская Россия) по совсем другому бюджетному обследованию на тот же период (см. Ковальченко И.Д. О буржуазном характере крестьянского хозяйства Европейской России в конце XIX - начале XX века // История СССР, 1983, N 5). Трудно сказать, связано ли это с простыми совпадениями или с пробивающейся везде неизменной закономерностью. В любом случае, сказанное повышает надежду на представленность в изучаемом бюджетном материале всех основных экономических типов хозяйств государственных крестьян региона в рассматриваемый период.

9. Ковальченко И.Д., Селунская Н.Б., Моисеенко Т.Л. Указ. соч., с.39.

10. Гусейнова А.С., Павловский Ю.Н., Устинов В.А. Указ.соч., с. 42

11. Ковальченко И.Д. Об изучении мелкотоварного уклада в России XIX в. // История СССР, 1962, N 1, с. 92.

12. Ленин В.И. Полн. собр. соч., т.16, с.204.

13. Ковальченко И.Д. Об изучении мелкотоварного уклада в России XIX в. // История СССР, 1962, N 1, с.92.

14. Рындзюнский П.Г. О мелкотоварном укладе в России XIX в. // История СССР, 1961, N 2, с.53.

15. Ленин В.И. Полн. собр. соч., т.16, с.201.

16. Ковальченко И.Д., Селунская Н.Б., Моисеенко Т.Л. Указ. соч., с.98.

17. Там же, с.88.

ВАРИАНТ СОЗДАНИЯ БАЗЫ ДАННЫХ О КРЕСТЬЯНСКИХ ХОЗЯЙСТВАХ УКРАИНЫ В ГОДЫ НЭПА

В.В.Подгаецкий, Ю.А.Святец

В статье рассматривается вариант создания проблемно-ориентированной базы данных о социально-экономической структуре крестьянских хозяйств Предстепья Украины по материалам весеннего выборочного опроса 1925 г., где число объектов задано самим источником, а число признаков представляется возможным сократить.

Аргументами в пользу сокращения размерности могут быть: 1) присутствие в наборе таких признаков, которые исследователь в момент создания базы данных склонен считать "случайными", т.е. не значимыми для описания искомых объектов; 2) отсутствие значений каких-то признаков у большинства объектов; 3) наличие такого значительного числа признаков, что, даже установив взаимосвязи между ними (например, с помощью корреляционного анализа), не возможно объяснить содержательный смысл этих связей.

В нашем случае для 7 округов (Александрийский, Екатеринославский, Зиновьевский, Краснодарский, Кременчугский, Первомайский и Полтавский), которые образовали собой Предстепье, были учтены хозяйства всех посевных групп, за исключением беспосевных. В базу данных "Крестьянские хозяйства Предстепья. 1925 год" (далее - БД "Предстепье") были включены данные для 77 признаков. Из-за отсутствия данных в БД не были включены такие признаки, как доля хозяйств без собственной удобной и собственной пахотной земли; поголовье коров в среднем на 100 хозяйств; доля полевого посева под табаком, кормовой свеклой, многолетними сеянными травами; доля хозяйств с 1 лошадей и 1 волом, с 2 лошадьми и 1 волом, с 3 волами. Наличие постоянных значений послужило основанием для того, чтобы не включать в БД следующие признаки: количество новых хозяйств, появившихся после раздела; число нанятых работников; доля посева под суржиком, бобовыми, сахарной свеклой, однолетними сеянными травами; доля хозяйств без коз. Малые значения с пропусками предполагают удаление из первичного набора таких признаков, как доля хозяйств, арендовавших только сеножати; поголовье рабочих коров в среднем на 100 хозяйств; размер посева на приусадебном участке, в том числе и картофеля. По той же причине были агрегированы показатели распределения хозяйств по количеству рабочего скота и коров. Например, вместо двух признаков - доля хозяйств с 3 коровами и доля хозяйств с 4 коровами - образован один новый - доля хозяйств с 3 и более коровами. С целью ограничения дублирования информации признаки

землепользования в расчете на едока не были включены в БД "Предстепье".

Традиционно задача сокращения размерности объектов решается путем сокращения описывающих их признаков. Однако это может привести к потере качества у изучаемых объектов (элементов), а следовательно, и у их совокупности в целом (т.е. у системы).

Отразить качественное своеобразие многомерных объектов может только система признаков, т.е. их эмерджентная совокупность. Последняя не является простой суммой качеств отдельных признаков, которые могут входить в качественное своеобразие объекта не целиком, а какой-то своей частью и влиять на его формирование с различной силой (мерой), установить которую возможно не *a priori*, а только в результате самого исследования. Признаки, считавшиеся историком случайными по причине "скрытости" информации, носителями которой они являются, будучи вписанными в результате структурного анализа их совокупности в систему, могут стать носителями знаний о ней. Другими словами, сокращение размерности путем уменьшения числа признаков вряд ли можно считать общим правилом при создании БД. Последнее заключение закономерно приводит к постановке кажущегося парадоксальным вопроса о том, нельзя ли уменьшить размерность объектов без фактического количества описывающих их признаков?

Результаты типологизации крестьянских хозяйств Украины, проведенной на основе анализа материалов весеннего выборочного опроса в 1925 г., дают возможность положительно ответить на поставленный вопрос. Методом, позволяющим сократить размерность объектов без потери их качества, в данном случае является метод главных компонент (МГК). Это происходит потому, что он учитывает значения всех признаков, используемых для описания объектов исследования (посевных групп крестьянских хозяйств). МГК позволяет не только сократить размерность объектов со 109 до числа учтенных главных компонент (ГК), но и выяснить причины качественного различия объектов и установить смысловое содержание ГК, выступающих в роли интегрированных признаков крестьянских хозяйств.

Обработка БД "Предстепье" методом главных компонент позволила выделить три из них (см. табл. 1), которые объясняли свыше 70% (72,3%) общей дисперсии объектов (посевных групп).

Таблица 1
Нагрузки признаков на компоненты для крестьянских хозяйств Предстепья (1925 г.)

I компонента (47,77%). Наличие комплекса производственных элементов

На одно хозяйство приходится:

+0,972 полевого посева
+0,966 пригодной земли в фактическом пользовании
+0,964 пригодной земли в собственном пользовании
+0,964 пахотной земли в собственном пользовании
+0,963 пахотной земли в фактическом пользовании

На 100 хозяйств приходится:

+0,96 всего голов скота
+0,951 коров
+0,931 плугов
+0,934 овец
+0,93 крупного рогатого скота
+0,928 членов семьи
+0,919 домашней птицы
+0,913 лошадей
+0,905 рабочих лошадей
+0,891 работников в рабочем возрасте
+0,89 свиней
+0,887 жаток
+0,84 мужчин
+0,749 буккеров

+0,917 доля хозяйств, обрабатывавших землю своим скотом (в %)
+0,912 доля хозяйств, обрабатывавших землю своим инвентарем (в %)
+0,875 доля хозяйств с двумя коровами (в %)
+0,854 доля полевого посева под овсом (в %)
+0,835 доля хозяйств, арендовавших только пахотную землю (в %)
+0,821 доля хозяйств с 3 и более головами рабочего скота (в %)

-0,907 доля хозяйств без рабочих лошадей (в %)
-0,903 доля хозяйств без рабочего скота (в %)
-0,899 доля хозяйств без лошадей (в %)
-0,899 доля хозяйств без овец (в %)
-0,872 доля хозяйств, обрабатывавших землю наемным инвентарем (в %)
-0,872 доля хозяйств без телег (в %)
-0,864 доля хозяйств, обрабатывавших землю наемным скотом (в %)
-0,85 доля хозяйств без орудий пахоты и посева (в %)
-0,823 доля хозяйств без свиней (в %)
-0,803 доля хозяйств без крупного рогатого скота (в %)
-0,862 доля хозяйств без коров (в %)

II компонента (12,29%). Промысловый характер хозяйства

+0,557 доля хозяйств без приусадебных и огородных посевов (в %)
+0,545 доля хозяйств без скота (в %)
+0,493 доля хозяйств, сдававших сенокосы в аренду (в %)
+0,459 доля хозяйств с 3 головами рабочего скота (в %)
+0,442 доля хозяйств без крупного рогатого скота (в %)
+0,439 доля хозяйств без орудий пахоты и посева (в %)
+0,417 доля хозяйств, обрабатывавших землю наемным скотом (в %)
+0,402 доля хозяйств, обрабатывавших землю наемным инвентарем (в %)

-0,845 доля хозяйств с 1 коровой (в %)
-0,815 доля хозяйств с 1 головой рабочего скота (в %)
-0,767 доля хозяйств, обрабатывавших землю супрягой скота (в %)
-0,743 доля хозяйств, обрабатывавших землю инвентарем смешанным способом (в %)
-0,739 доля хозяйств, обрабатывавших землю супрягой инвентаря (в %)

III компонента (8,67%). Рыночная ориентация земледелия

+0,687 доля полевого посева под кукурузой (в %)
+0,675 доля полевого посева под ячменем (в %)
+0,58 доля полевого посева под масличными культурами (в %)
+0,51 сдано десятин пахотной земли одним хозяйством, сдававшим ее
+0,504 доля полевого посева под озимой пшеницей (в %)

-0,754 доля полевого посева под яровой пшеницей (в %)
-0,502 доля полевого посева под гречкой (в %)
-0,466 на 100 хозяйств приходилось рабочих волов
-0,465 доля полевого посева под озимой рожью (в %)
-0,41 доля полевого посева под бахчевыми и огородными культурами (в %)

Первая компонента (51,33%) имеет положительную связь с признаками, характеризующими потенциальную производственную мощность хозяйств, а именно: наличие собственной рабочей силы, размеры землепользования и аренды земли, наличие продуктивного и тяглового скота, домашней птицы, а также сельскохозяйственного инвентаря и транспортных средств (телег). Частично компонента касается производственных отношений. Речь идет об обработке земли собственными средствами, а также о сельскохозяйственных культурах, являющихся индикатором определенного развития и организации крестьянских хозяйств, озимой пшенице и овсе.

Наоборот, отрицательная связь компоненты наблюдается с признаками, характеризующими хозяйства без собственного скота, орудий труда, вследствие чего они вынуждены были нанимать скот и инвентарь для обработки земли, которую засеивали обычно рожью, а используемую часть которой сдавали в аренду, занимаясь при этом сторонними заработками. Таким образом, первая компонента определяет наличие комплекса производственных элементов как условия для определенной организации и направления деятельности (специализации) крестьянских хозяйств.

Вторая компонента (12,29%) положительно связана с признаками, которые характеризуют две категории крестьянских хозяйств. К первой из них относятся хозяйства, имеющие довольно низкий уровень обеспеченности скотом (в том числе и рабочим), а также крупным рогатым скотом (в том числе и коровами). Отсутствие орудий для пахоты и посева, транспортных средств, телег приводило к тому, что они обрабатывали землю наемным живым и мертвым инвентарем, и засеивали ее, главным образом, рожью и ячменем. В другом случае эти хозяйства сдавали пахотную землю и сенокосы в аренду, а поскольку ренты при этом не хватало для удовлетворения собственных потребностей, вынуждены были отчуждать рабочую силу на заработки.

Вторая категория крестьянских хозяйств имела более благоприятные условия, чем предыдущая. Эти хозяйства обычно держали не менее трех голов рабочего скота или трех коров. К тому же значительная часть их обладала сложными машинами для сбора (жатки) и переработки (обмолота) урожая.

Негативная связь компоненты наблюдается с признаками, характеризующими хозяйства с ограниченным поголовьем рабочего (1-2 головы на двор) и продуктивного (1 корова) скота. Они обрабатывали землю преимущественно супрягой и смешанным способом или сдавали определенную часть земли в аренду, реже арендовали. При этом доля полевого посева от фактического пользования пригодной землей была довольно высокой, то есть хозяйства прикладывали все усилия к тому, чтобы максимально использовать землю. В структуре полевого посева преобладают такие культуры, как ячмень (на корм лошадям), огородные и бахчевые культуры.

Таким образом, вторая компонента характеризует хозяйства, ориентировавшиеся на получение главного дохода вне собственного сельскохозяйственного производства (обслуживание сельского хозяйства), то есть это компонента "промысловости".

Третья компонента (8,67%) положительно связана с признаками, свидетельствующими об отношениях аренды пахотной земли и сенокосов. Характерно, что арендованная земля использовалась исключительно под посев рыночных культур - кукурузы, ячменя, озимой пшеницы, масличных растений. В животноводстве приоритет отдается свиноводству. Несмотря на то что среднее поголовье лошадей здесь сравнительно высокое и в

этой группе сосредоточено определенное количество сложных орудий труда (буккеры), обработка земли осуществлялась обычно смешанным способом использования тягла и инвентаря.

Негативная связь компоненты наблюдается с признаками, характеризующими интенсивное земледелие хозяйств, - посевы яровой пшеницы, озимой ржи, гречихи, бахчевых и огородных культур. Обработка земли осуществлялась, как правило, рабочими волами с использованием наемного труда.

Вышесказанное позволяет сделать вывод, что третья компонента характеризует рыночную направленность земледелия.

Выделив ГК, возможно провести типологизацию крестьянских хозяйств (с помощью кластер-анализа) исходя из того, что в роли признаков, характеризующих объекты, будут выступать ГК, а в роли их значений - индексы ГК для каждого из объектов (посевных групп крестьянских хозяйств). В результате типологизации были получены 6 типов социально-экономических структур крестьянских хозяйств Предстепья:

1. Маломощные потребляющие хозяйства со сторонними заработками (4,3%).
2. Маломощные хозяйства с рыночной ориентацией земледелия (19,7%).
3. Крепкие хозяйства с рыночной ориентацией земледелия (40,6%).
4. Крепкие хозяйства исключительно рыночноориентированного направления земледелия (23,2%).
5. Мощные хозяйства без определенной специализации (7,1%).
6. Мощные хозяйства комплексного характера сельскохозяйственного производства с торгово-промышленными предприятиями (0,9%).

Преимущество данного варианта типологизации заключается в том, что фактически выявление типов происходит не в пространстве 109 признаков, роль и значение многих из которых для характеристики исследуемых объектов могут быть наперед не известны, а в пространстве достаточно определенных признаков - компонент, качественное своеобразие и сущностно-содержательный смысл которых нам уже известны. Сам процесс типологизации в этом случае возможно определить как разделение совокупности объектов на типы в зависимости от того, в какой мере в каждом из них проявляются качества ГК, т.е. факторных признаков. Результаты типологизации крестьянских хозяйств, проведенной обоими вариантами, - сопоставимы.

Резюмируя, содержательный смысл произведенных операций и их результатов возможно изложить в следующем виде: 1) с помощью МКГ проведена структурализация генеральной совокупности признаков, характеризующих социально-экономическое устройство крестьянских хозяйств, 2) в роли элементов ее структуры выступают эмерджентные совокупности признаков, качественное своеобразие которых выражается через качественное своеобразие ГК, 3) набор ГК возможно рассматривать как репрезентативную выборку качеств признаков, элементами которой

являются не сами признаки, а интегрировавшие их факторы (ГК), оказывающие решающее влияние на формирование качественного своеобразия системы изучаемых объектов. В итоге сокращение размерности происходит без уменьшения числа признаков, т.е. без потери качества, 4) при решении задач типологизации объектов создание реляционных БД с помощью МГК представляется обоснованным и необходимым шагом.

Алгоритм решения проблемы:

1. Исходная позиция - первичная БД в виде электронной таблицы, где n - число объектов, а m - число признаков (109).

2. Обработка первичной БД с помощью МГК. Результат - вторичная БД, где n - число объектов, p - число учтенных главных компонент.

3. Представление вторичной БД в виде k (по числу главных компонент), качественно-ориентированных на содержательный смысл каждой из учтенных ГК. Результат формализуется в виде k электронных таблиц, где n - число объектов, $n(i)$ - число признаков, связанных с данной компонентой ($i=1, k$).

4. В ходе решения задачи социально-экономической типологизации объектов для различных природно-географических регионов описанная выше операция производится для каждого из них. В результате имеем реляционные БД, объединенные в банки данных (БД), число которых равно количеству регионов, в пределах каждого из которых осуществляется типологизация объектов (крестьянских хозяйств).

5. Являясь эмерджентными совокупностями, компоненты, выделенные для различных регионов, могут представить схожие качества, сопоставимые между собой. Это позволяет сопоставлять и выявлять тенденции в качественном состоянии типов из различных регионов, т.е. определять их как качественно сопоставимые, даже когда они характеризуются различными наборами признаков, или не сопоставимыми по количественному выражению значений этих признаков. Методологическими посылками, исходя из которых возможно осуществить данную операцию, являются: 1) многомерность исследуемых объектов; 2) представление о качестве совокупностей (типов) объектов как об эмерджентных совокупностях признаков.

IV. КОМПЬЮТЕР В ОБРАЗОВАНИИ ИСТОРИКА

МНОГОМЕРНОСТЬ ПРЕДСТАВЛЕНИЯ ЗНАНИЯ В ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЯХ ОБУЧЕНИЯ

Н. И. Миницкий

В современных информационных технологиях обучения вполне отчетливо прослеживается тенденция количественного роста различных методик и форм представления знания. Меняется характер отношений между традиционными методиками и компьютерными технологиями обучения. Непримируемость и взаимное неприятие сменяются стремлением к сближению. Стоит вспомнить лишь один из моментов - идею о замене учителя компьютером - чтобы представить себе остроту прежних коллизий. Не "вместо", а "вместе и для" - такова современная позиция в этом вопросе. Теперь для технологии обучения все более естественной становится связь между дидактикой, когнитивными науками и техническими средствами. Их взаимодействие приобретает диалектический характер.

Первоначальное акцентирование внимания на технологических аспектах компьютерного обучения нашло отражение в самом понятии "компьютерные технологии". В дидактику это понятие внесло жесткость алгоритма обучения и угрозу догматизации мышления. В современном понимании информационных технологий более явно обнаруживается связь между методологическими и техническими средствами /1/. Информационные технологии воспринимаются как совокупность используемых в учебном процессе новых методов и средств сбора, хранения, обработки, передачи и представления научных знаний на базе электронно-вычислительной техники /2/.

В дидактической литературе довольно четко определена задача разработки методик учебно-познавательной деятельности на основе методов научного познания /3/. Это одно из центральных направлений развивающего обучения. Аналогична ситуация и в методологических исследованиях, где отмечается необходимость разработки процедур, переводящих философскую и общеметодологическую теорию в практику конкретно-научного действия /4/. Функцию конкретных методик - служить

связующим звеном теории и практики обучения – взяли на себя информационные технологии. Их оценивают как основу педагогики завтрашнего дня.

Психолого-дидактические приемы обучения, предметное содержание и технические средства составляют обучающую среду /5/. Одним из ее важнейших качеств является многомерность. Понятие многомерности довольно прочно утвердилось в одном из направлений информационных технологий – количественных методах исторического исследования /6/. На наш взгляд, имеется достаточно оснований для распространения этого понятия на всю обучающую среду.

В предлагаемой вниманию читателя статье делается попытка выяснить в какой степени идея многомерности отражена в информационных технологиях, когнитивных моделях обучения и формах представления знания.

Какие же информационные технологии в большей степени соответствуют идее многомерности? Для ответа на этот вопрос обратимся первоначально к визуальному мышлению и соответствующим образам изобразительного искусства. В частности, линейная система первых автоматизированных систем обучения ассоциируется с ритмичкой орнамента первобытной керамики. Иерархически ветвящаяся система семантических сетей и фреймов как бы оживает в древневосточных рельефах, где все объединено единым замыслом сюжета, а изображения расположены горизонтальными подчиненными рядами (см. знаменитую "Стеллу коршунов"). Непредсказуемость и мягкость форм гипертекстовой структуры соответствует и самым смелым современным замыслам конструкций кубизма. Здесь связи элементов системы не намечены заранее, а носят вероятностный характер. Гипертекст и кубизм – продукт контрастов и ярчайшее проявление многомерности. Система ключевых слов – основа базы данных современных информационных технологий, в том числе и гипертекста, – имеет логическое сходство с древнекитайскими текстами, где существовали категориальные матрицы познания.

Современные исследователи в области информационных технологий находят гипертекстовую структуру весьма гибким средством обработки информации в различных интеллектуальных сферах человеческой деятельности: науке и технике, истории и политике, в военном деле и образовании, словом, там, где необходима обработка больших массивов информации.

Практика применения гипертекстовой технологии в обучении позволяет выделить следующие, на наш взгляд наиболее существенные дидактические и познавательные признаки этой системы. Самым важным ее качеством является большая, чем у других информационных технологий, возможность свободы выбора вариантов решений и творческой интерпретации. Попутно отметим еще одно примечательное явление: чем существеннее свойство, тем оно имеет большую историческую традицию. Стоит вспомнить

особенности раннего шумерийского письма на Древнем Востоке: письменный знак мог передавать не только одно, но и определенное количество ассоциативно связанных понятий /7/. Исполнитель письменного текста, не имея возможности дословного воспроизведения, при чтении опирался на ключевые понятия и свободно его интерпретировал. Не линейное чтение информации является одной из важнейших особенностей и гипертекстовой технологии. Внутренняя связь информации смысловых единиц гипертекста обеспечивается формально-логическими структурами, а внешняя – произвольна и чаще всего ассоциативно. Главное – это внешние связи.

Подобная особенность гипертекста порождает возможность свободы выбора информации и творческой интерпретации фактических данных. Историк данную ситуацию может реализовать не только в выборе разных взглядов, что значимо само по себе, но и во всестороннем анализе источника. Текст источника в гипертекстовой обработке, более чем в любой другой форме представления знания, выглядит как рефлексия исторической действительности. Во многом, благодаря линейной системе заранее сформулированных вопросов, в учебных методических указаниях старого типа текст источника превращался в иллюстрацию идеологического заказа. Применение заранее изготовленных теоретических или методологических установок вносит в исследования и обучение большую или меньшую долю субъективизма. В таком случае наиболее объективным средством обработки источников оказываются количественные и формально-логические методы. Их применение позволяет наиболее полно отразить содержание источника и извлечь самую разностороннюю информацию. Использование компьютерных технологий позволяет более эффективно решать задачи обработки источников в учебно-познавательных и исследовательских целях.

Средством познавательной деятельности, кроме математических и формально-логических методов, могут быть также когнитивные модели, построенные на основе самого исторического знания. Исходя из современного состояния разработки методологии познания и общедидактических требований, попытаемся сформулировать некоторые общие требования к когнитивным компьютерным моделям.

1. Для создания компьютерных моделей необходимо провести структурирование учебного материала и разработать концептуальные модели. Они представляют собой единство категориальной структуры и предметного содержания /8/. Категориальная структура должна включать категориальное знание философского, общенаучного и частнонаучного уровней. Категориальное знание, дополненное фактическим материалом, составляет основу современной информационной технологии фреймов.

2. Модели должны обладать свойством операциональности, т.е. иметь описание последовательности познавательных процедур. Операциональность находит конкретное выражение в свертывании и развертывании информации по концептуальным моделям /9/. Процедура свертывания информации

практически реализуется в системе предполагаемых ответов, находящихся в памяти компьютера. Наибольшая степень сжатия информации достигается при использовании формы свободно-конструируемого ответа, когда жестко фиксируются лишь ключевые слова, а второстепенные связи и признаки интерпретируются свободно.

3. Одно из главных требований к моделям - разнообразие их форм и объема содержания. Существующие формы модельного представления знаний вполне отвечают подобным требованиям. Главные виды моделей - репрезентации формально-логических и лингвистических структур /10/, понятийные формы знания /11/, имитационные и концептуальные модели /12/, универсальные объяснительные схемы /13/, фреймы /14/. Разнообразие исходных оснований построения моделей дает возможность видеть в этом факте выражение многомерности.

Итак, внутренняя структура знания предстает перед нами преимущественно в форме многомерных моделей, главными признаками которых являются категориальность, концептуальность и операциональность знания.

Теперь обратимся к следующему важному аспекту выражения многомерности - внешним формам представления знания. В качестве исходного момента для рассуждения отметим, что учебная информация предстает перед обучаемым преимущественно в двух формах: вербальной и символической. Их дополнение музыкой, цветовыми и другими эффектами, а также другими техническими средствами позволяет говорить о новой обучающей среде - мультимедиа и соответствующей ей технологии - гипермедиа. Возникает вопрос: в какой мере эта среда соответствует идее многомерности.

Для того, чтобы на него ответить обратимся к конкретному рассмотрению текстово-символьных форм представления знания. Для историка основным средством учебной информации служит текст, представленный в учебной, художественной и публицистической литературе. Символьная часть информации включает рисунки, графики, схемы, диаграммы и т.д. Значение символической информации выходит далеко за пределы чисто иллюстративного характера. Стоит вспомнить те надежды, которые возлагались в философской учебной литературе на различного вида схемы. Ставилась цель способствовать восприятию абстрактных категорий философского знания. Однако отсутствие связей и моделей с конкретно-философским содержанием обучения и абсолютизация графических форм стали причиной падения их популярности.

Историкам удалось избежать крайностей формализации и, более того, благодаря мастерству педагога-новатора Г.Ф. Шаталова вновь возрос интерес к подобного рода средствам наглядности. Отдавая должное индивидуальному творчеству, хотелось бы выделить необходимость отражения мировой символики в символической части представления знания. Теоретической основой содержания символического представления знания

могут послужить работы А.Ф. Лосева /15/. А современные теоретические изыскания в области визуального мышления /16/ станут исходным моментом для создания практических рекомендаций по применению визуальных средств в обучении. Дидактическая сторона проблемы применения знаков и символов в обучении достаточно полно изложено Н.Г. Салминой /17/.

Дополнительный импульс к применению символов в обучении дало возникновение компьютерной графики. Порой ее называют "когнитивной графикой" или даже "картинной логикой" и считают, что в недалеком будущем состоится замена текстовых интеллектуальных систем текстово-зрительными. Признавая значимость этой тенденции предлагаем общую модель вербального и символического представления информации на основе применения компьютерной графики. Подобная форма представления знания может быть названа "когнитивной картой". Она выгодно отличается от прежних схем тем, что объединяет абстрактный материал общих понятий с конкретно-образным мышлением. На основе информационных технологий – фреймов и гипертекстов – в обе части "когнитивной карты" вводится конкретно исторический материал. Последующие дополнения понятийной модели символическим материалом создает дополнительные возможности взаимодействия абстрактного и конкретного мышления.

Компьютерный вариант "когнитивной карты" может стать средством компромисса между сторонниками введения в компьютер всего текста учебника и теми, кто придерживается фрагментарного представления материала /18/. Обе теории имеют равноценные и крупные недостатки. Введение всего текста слишком громоздко и дорогостояще, а фрагментарное не создает целостного видения предмета. Компьютерная "когнитивная карта" позволяет создать системное представление о предмете в его абстрактных и конкретных понятиях.

"Когнитивная карта" в компьютерном варианте может быть использована как в режиме обучения, так и контроля. В режиме обучения необходимо ее дополнение фрагментами текстов и разного рода практическими заданиями. В целом "когнитивная карта" является средством, синтезирующим различные формы представления знания. Компьютерные средства делают эти формы более эффективными за счет постоянного пополнения баз данных и баз знаний, а также возможности быстрого поиска и создания новых комбинаций знания. Задача новых информационных технологий, трансформированных в технологию обучения, состоит прежде всего в решении проблемы порождения знаний, т.е. в обнаружении, понимании и изобретении смысловых структур. В дидактическом осмыслении – это задача научить учиться.

Что же служит препятствием на пути к быстрейшему и эффективному применению компьютерных технологий в обучении? Основная причина лежит в противоречиях между существующей теоретико-методологической средой информационных технологий и содержанием обучения. Очевидно, предпосылкой преодоления этого противоречия является разработка

познавательных средств и форм представления материала в учебной литературе. Оставляя в стороне теоретико-содержательную часть учебника обратимся к его инструментально-практической части. С точки зрения современной теории учебник должен содержать изложение основных методов познания, а так же прикладных операций исторического исследования. Разнообразие форм представления материала способствовало бы сближению обучающей среды с образами мира истории. Все это позволило бы студенту не только повысить методологическую культуру, но и пользоваться современными технологиями в креативном получении знаний.

ПРИМЕЧАНИЯ

1. P.N. Johnson-Laird. The Computer and the Mind. An Introduction to Cognitive science. Cambridge 1988.

2. Новые информационные технологии в учебном процессе: организационный опыт Московского университета // Информационный Бюллетень Комиссии по применению математических методов в ЭВМ в исторических исследованиях при отделении истории РАН, 1991, N3, с. 29.

3. Давыдов В.В. Проблемы развивающего обучения. М., 1986, С. 83-86, 91, 205; Марев И. Методологические основы дидактики. М., 1987, С. 153; Гершунский Б.С. Компьютеризация в сфере образования. М., 1987, с. 118,185.

4. Ворожцов В.П. Роль методологии в повышении эффективности научных исследований // Методология в сфере теории и практики. Новосибирск, 1988, с. 80.

5. О понятии обучающей среды см.: Вербицкий А., Агапова О., Ушаков А. Сотворение образа мира // Вестник высшей школы. 1991, N7, с. 26.

6. Бородкин Л.И. Многомерный статистический анализ в исторических исследованиях. М., 1986.; Ковальченко И.Д. Методы исторического исследования. М., 1987, с. 416-435.

7. Дьяконов И.М. Шумерийский язык // Языки Азии и Африки. III. Языки древней Передней Азии. М., 1979, с.9.

8. Свириденко Д.И. Предпосылки логического подхода к программированию // Философские основания научной теории. Новосибирск, 1985, с. 171.

9. Блюменау Д.И. Проблема свертывания научной информации. Л., 1982.

10. Вартофский М. Модели. Репрезентация и научное понимание. М., 1988, с. 10, 67.

11. Соловьева Е.А. Теория понятийных знаний. Киев, 1990.

12. Устинов В.А., Кузищин В.И., Павловский Ю.С., Гусейнов А.С. Опыт имитационного моделирования историко-социального процесса // Вопросы истории. 1986, N4; Шенк Р.К. Обработка концептуальной информации. М., 1980, с.159.

13. Сергеев В.М. Искусственный интеллект: опыт философского осмысления // Будущее искусственного интеллекта. М., 1991, с. 236.
14. Минский М. Фреймы для представления знаний. М., 1979.
Гончаренко В.В., Шингарева Е.А. Фреймы для распознавания смысла текста. Кишинев, 1984.
15. Лосев А.Ф. Проблема символа и реалистическое искусство. М., 1976.
16. Жуковский В.И., Пивоваров Д.В., Рахматулин Р.Ю. Визуальное мышление в структуре научного познания. Красноярск, 1988.
17. Салмина Н.Г. Знак и символ в обучении. М., 1988.
18. О наличии подобных взглядов см.: Розенберг Н.М. Компьютер и учебник // Советская педагогика. 1988, №6, С. 31.

КОМПЬЮТЕРНАЯ ТЕХНОЛОГИЯ ОБУЧЕНИЯ ИСТОРИИ
В БЕЛОРУССКОМ ГОСУДАРСТВЕННОМ УНИВЕРСИТЕТЕ: ПУТИ РЕАЛИЗАЦИИ

Е.Н.Балыкина, В.Н.Сидорцов

Использование компьютерной технологии обучения (КТО) в преподавании истории зависит в первую очередь от наличия машинной техники соответствующего класса и программного обеспечения. Исходя из имеющихся технических возможностей (оснащенность компьютерного класса несколькими ПЭВМ ЕС-1841) и обобщая свой опыт применения компьютеров, можно сделать вывод, что наиболее целесообразными формами обучения на базе ЭВМ являются следующие:

- 1) вопросно-выборочная;
- 2) обучающе-игровая;
- 3) вопросно-ответная (свободно-контролируемая).

Педагогу, начинающему применять КТО, лучше всего обратиться к текстовому опроснику (1-й форме). При этом он может осуществлять контроль различными способами:

- последовательная подача вопросов от 1 до N из банка заданий, например в 30 вопросов;
- опрос посредством датчика случайных расстановок вопросов;
- контроль того же количества случайно подаваемых вопросов, однако из банка заданий, число которых значительно увеличено, например в 5 раз (150 вопросов);
- предъявление для узнавания ответов (оптимально 5) по каждому из вопросов и каждому из студентов (естественно, что правильный ответ будет занимать разные позиции);
- оптимальный вариант, когда используются одновременно последние два способа, причем формируется мини-банк из 10 вопросов (9 неверных и 1 верный), тогда каждый из обучаемых получит случайным образом расположены 1 верный и 4 из 9 неверных ответов.

Особый интерес представляет обучающе-игровая форма. Обучающие игры, в противоположность развлекательным, становятся неотъемлемым элементом компьютерной технологии обучения. И это не случайно, ибо с их помощью можно решать большой круг педагогических задач, поскольку игры обладают рядом психолого-педагогических преимуществ и оказывают положительное влияние на обучаемого. В сферу применения игр входит и гуманитарное образование, в частности историческое. Опыт, накопленный на кафедре источниковедения и музеологии исторического факультета БГУ, однозначно подтверждает необходимость данной игровой компоненты и ее более широкого применения.

Элементы игры целесообразно использовать в той или иной форме во всех компьютерных программах обучающего характера. Однако необходимы и специальные (отдельные) игровые программы. В зависимости от состава аудитории, содержания изучаемой темы и поставленной дидактической задачи они могут быть представлены двумя разрядами: 1) простым или упрощенным (стрельба датами по условной мишени, ходьба по лабиринту истории, узнавание исторических объектов); 2) сложным или интеллектуальным (исторические головоломки, шашки, рулетка и т.д.).

Работа с отдельными обучающими играми по истории вылилась в создание сериала игр, объединенным одним и тем же видом, но позволяющим неоднократно возвращаться к нему в различной игровой форме. При этом учтены следующие требования: соответствие предметного содержания игр современному уровню научного знания, наиболее эффективное использование компьютера, возможность тестирования и свободного выбора различных фрагментов и уровней сложности, контроль и анализ знаний играющих. На этой основе разработана серия обучающих головоломок по истории мировой и отечественной культуры. Показательной в этой серии является пакет "Головоломка", состоящий из кроссворда, кроссчайнворда и шарады. Ее авторы учитывали уровень познавательной готовности студентов к выполнению поставленной задачи, исходили не только из требований класса релевантной избыточной информации, а и других психологических особенностей обучаемых.

Каждая из программ серии предназначена для обучения и контроля знаний и содержит такие элементы, как рисунки (кроссворд, чайнворд, шарада, ребус, метаграмма и др.), перечень вопросов, предметные подсказки трех уровней (по внешним признакам искомого объекта, по качественной характеристике, по отдельным буквам), помощь пользователю программы, информация о работе в режиме "обучение", неограниченного во времени, и о происхождении контроля, ограниченного во времени. Обучаемый решает задачи самостоятельно, а в случае затруднения - с помощью ПЭВМ. Работа пользователя оценивается в определенной системе баллов.

Игра организуется таким образом, что обучаемый вначале набирает темп, отвечает на все известные ему вопросы, а затем, предельно мобилизуя свои способности, получая помощь всех уровней в поисках ответа на те вопросы, по которым допустил ошибки или на которые вообще не ответил, приходит к правильной мысли самостоятельно либо по наводящим признакам заданного объекта. Познавательная деятельность стимулируется также своим режимом занятия (тренаж или контроль), а в каждом из них - оценками ответов и учетом времени, затрачиваемым на работу, характером приза, выдаваемого за решение задачи-головоломки, и др.

Существенное значение при разработке обучающих игр имеет:

1/ оптимальное сочетание традиционных правил составления игрово-головоломок с учетом введения в оборот не более 20 слов и возможностей компьютера (ограниченность объемов экрана, но использование диалогового режима, динамической графики и др.);

2/ содержательный отбор материала для программирования собственной "энциклопедии-матрицы" - базы данных (например, по истории культуры - на одной оси в системе координат обозначаются все области культуры, на другой - буквы в алфавитном порядке, на третьей - хронология);

3/ разработка вариантов той же программы (серии программ) с целью создания у каждого из обучаемых иллюзий неповторяемости его игры путем изменения вида задачи-головоломки (кроссворд, чайнворд, шарада...), формы самого рисунка в рамках того или иного вида (круглая, прямоугольная, похожая на улитку и др.), фоновых элементов (образное изображение определенного предмета из числа основных, составляющих тему, например, для чайнворда - река или дорога и т.д.) или сюжетных элементов (изображение динамики процессов, характерных для рассматриваемого объекта, имитация действий участника игры). А следовательно, изменения и части вопросов, однако с таким расчетом, чтобы основополагающая их часть присутствовала в каждом варианте, а общее количество в полной мере охватывало важнейшие события или явления компьютеризируемой темы.

С учетом опыта было разработано компьютерное сопровождение к курсу истории Беларуси (1917-1993 гг.) для 9-го класса средней школы. Он включает три больших программы: - тема "Противостояние большевиков и их противников: 1917-й - середина 20-х гг." реализована в виде "исторических" шашек, вторая - "Беларусь в годы тоталитаризма" - трех компьютерных лабиринтов и третья - "Беларусь в годы второй мировой войны" - серии кроссвордов (программы проходят апробирование в школах Минска).

В соответствии с идеей многократного обращения к системе игр, связанных единой сюжетной линией, создавалась обучающая игра по всемирной истории под условным названием "Кругосветное путешествие". Первым шагом в ее реализации являлась разработка блока "Российский трон". Однако его апробирование показывает, что подобный сериал представляет ценность лишь с точки зрения приобретения и закрепления обобщенных знаний сравнительного характера.

В интересах более глубокой интеллектуализации работы с ПЭВМ целесообразно использовать существующие компьютерные игры, создавая на их базе программы обучающего характера. Например, по образцу настольной игры "Шанс" разработана компьютерная обучающая игра "Деятельность монополий в годы мирового экономического кризиса (1929-1933)". Обращение к ее банку данных, в котором содержатся необходимые сведения о монополиях, и к разнообразному материалу справочного характера позволяет играющему успешно справляться с ролью владельца

монополий. Программа послужила основой для разработки новой игровой серии.

Вместе с тем критический анализ программы привел к мысли о необходимости создания учебных баз данных с игровым элементом. Подобные базы разрабатываются по актуальным темам отечественной истории ("Республика Беларусь: национальное движение", "Новые подходы к истории Беларуси: анализ новейшей литературы" и др.). Постоянно пополняясь новым историческим материалом, они также могут стать игровыми сериалами.

С целью автоматизации проектирования игровых обучающих программ создана специальная система "Проект". Она состоит из двух основных комплексов (программа проектирования сценариев и интерпретатор автоматизированных учебных курсов), управляемых диалоговым программным монитором. Система дает возможность пользователю, не являющемуся профессиональным программистом, разрабатывать автоматизированные курсы и применять их по классу ПЭВМ - IBM PC/XT.

Свободно-конструированная форма предполагает работу обучаемых в четырех вариантах - входное и выходное тестирование, ознакомление с "теорией", обучение с элементами тренажа, контроль. С помощью входного теста выявляется уровень подготовки студентов к работе с программой, в зависимости от чего ему предлагается подключиться к теоретической части, если он показывал крайне слабые знания, либо войти в режим "обучение", при средних показателях, либо перейти сразу к контролю, при отличных результатах.

Теоретическая часть представляет собой квинтэссенцию учебного материала, изложенного в наиболее удобном для восприятия, запоминания и воспроизведения виде. При ее формировании используется система опорных зон, т.е. знаков и символов, включая схематические рисунки, ключевые слова, мнемонические приемы, короткие тезисы. "Теория", с одной стороны, читается как единый связанный текст, с другой - воспринимается как сумма элементов информации, емкость которых составляет от 1/2 до 4 экранов дисплея, что вполне достаточно для освещения отдельного вопроса или их группы. Порция выдается на экран по запросу студентов в режиме "обучение". Каждая из них существует и как верный развернутый ответ на поставленный вопрос, и как "эмоциональное пятно", способствовавшее лучшему усвоению вопроса /яркий фрагмент летописи, исторический анекдот, увлекательные строчки исторического романа/, и как вывод, содержащий только ключевые слова.

Режим "обучение" включает элемент тренажа, алгоритм прохождения которого заставляет обучаемого упорно искать верные ответы на все вопросы. При этом он может воспользоваться "теорией", получить справку, помощь или подсказку. При этом предусматривается полная диагностика типовых ошибок и неполных ответов.

Компьютерную технологию обучения наиболее целесообразно использовать в следующих случаях:

1/ при самостоятельной работе студентов без участия преподавателя или при контроле знаний студентов по темам, не представляющим больших трудностей;

2/ при необходимости тренажа с целью повышения знаний отстающих студентов до уровня, необходимого, чтобы воспринимать лекции или укладываться во время, отводимое для подготовки к практическим /семинарским/ занятиям.

Опыт показывает, что компьютерные технологии следует сочетать с традиционными методами обучения. При этом необходим тщательный учет применимости тех или иных в каждом конкретном случае. Комплексность подхода предполагает не только такой учет: количественный, порядковый, соответствующий содержанию изучаемого материала. Немаловажное значение имеет внутренняя взаимосвязь всех методов обучения, которая обнаруживается при сохранении и отслеживании истории обучения каждого студента от его первого контакта с преподавателем и ПЭВМ до самостоятельной защиты диплома.

Комплексный пакет педагогических программных средств должен содержать такие блоки, как: а/ обучение пользованию компьютером с максимально дружественным интерфейсом, погружающим обучаемого в состояние повышенного психологического комфорта, что особенно важно для гуманитария; б/ психологическое тестирование, которое позволяет выявить психофизиологические характеристики, влияющие на обучение; в/ адаптация к обучаемым в зависимости от их психофизиологических характеристик; г/ релаксопедия; д/ обучение-контроль; е/ рубежный.

Психологическое тестирование позволяет выявить наряду с другими такую характеристику, как тревожность, которая оказывает довольно значительное и часто отрицательное влияние на ход обучения, в частности увеличивает время работы обучаемого с компьютерной программой. Найденный метод адаптации обучаемых состоит в формировании кадров обратной связи (КОС) в зависимости от уровня их тревожности.

Под КОС понимается реакция АУК на сообщение обучаемого того или иного характера: верный, неполный, непредусмотренный или заведомо неверный, повторно непредусмотренный ответы, типовая ошибка, запросы помощи, подсказки, справки. Кадры формируются блоком адаптации и подключаются к блоку обучения-контроля. Каждый кадр состоит из констатирующей, адресной, эмоциональной и итоговой частей, где содержится определенная реакция на тот или иной уровень тревожности. При высокой тревожности КОС направлены на то, чтобы поддержать обучаемого, вселить в него уверенность, при низкой - "завести" на более результативную работу. Реплики выбираются случайным образом. Их банки формируются с учетом ограниченной емкости экрана дисплея, напряжения зрительного анализатора, экономии компьютерного времени и

целостности всего КОС. Их использование создает необходимое настроение для успешной работы с компьютером.

Другим немаловажным элементом компьютерной технологии является релаксопедия, позволяющая снять усталость пользователю ПК. Блок релаксопедии представляет собой совокупность тематически однородных и небольших банков шуток, анекдотов, курьезных случаев из жизни людей. С помощью датчика случайных расстановок выбирается как сама тема, так и кадры шуток. При создании такого блока необходимо придерживаться правила "золотой середины". Десять экранов с 2-4 эпизодами вполне достаточно, чтобы создать обстановку, располагающую к результативной работе, повысить работоспособность.

Многолетний опыт создания и внедрения в учебный процесс КТО привел к выявлению наиболее рациональных путей ее использования.

СВЕДЕНИЯ ОБ АВТОРАХ

М.Т.Абдулганеев
656099 Россия Барнаул, ул.Димитрова, д.66
Алтайский государственный университет
исторический факультет

к.и.н. В.Л.Акимов
430000 Россия, Саранск,
ул.Большевиктская, д.68
Мордовский государственный университет
исторический факультет

к.и.н. Н.М.Арсентьев
430000 Россия, Саранск,
ул.Большевиктская, д.68
Мордовский государственный университет
исторический факультет

Е.Н. Балыкина
220080 Беларусь, Минск, пр-т Франциска Скорины, д.4
Белорусский государственный университет
исторический факультет

д.и.н. Ю.П.Бокарев
117036 Россия, Москва, ул.Дм. Ульянова, д.19
Институт Российской истории РАН

д.и.н. Л.И.Бородкин
119899 Россия, Москва, Ленинские горы
Московский государственный университет
исторический факультет

А.В.Быстров
Россия, Москва
ГИВЦ г.Москвы
Дом. адрес: 121019, г.Москва, просп.Калинина, д. 24 кв. 19

к.и.н. В.Н.Владимиров
656099 Россия, Барнаул, ул.Димитрова, д.66
Алтайский государственный университет

исторический факультет

к.и.н. И.М.Гарскова
119899 Россия, Москва, Ленинские горы
Московский государственный университет
исторический факультет

к.и.н. Д.А.Гутнов
117036 Россия, Москва, ул.Дм. Ульянова, д.19
Институт Российской истории РАН

к.и.н. Е.В.Злобин
103107 Россия, Москва, К-107, Б.Садовая, д.14
Гуманитарная Академия Вооруженных Сил
Российской Федерации

Д.Х.Ибрагимова
119899 Россия, Москва, Ленинские горы
Московский государственный университет
исторический факультет

к.и.н. Н.И.Миницкий
Белорусский государственный экономический университет
кафедра филологии
Дом. адрес: 220028, Беларусь, Минск, ул.Маяковского, 152, кв.79

к.и.н. В.Л.Носевич
220116 Беларусь, Минск, ул.Коллекторная, д.10
Комитет по архивам и делопроизводству при СМ РБ

В.А.Перевертень
Россия, Москва, Российский государственный
гуманитарный университет
Дом. адрес: 127644, Россия, Москва, ул. Лобненская, 12-1-421

д.и.н. В.В.Подгаецкий
320625 Украина, Днепропетровск, пр. Гагарина, д.72
Днепропетровский государственный университет
исторический факультет

д.и.н. И.Р.Рафи-Заде
Бакинский государственный университет
исторический факультет
Дом. адрес: 370010, Азербайджан, Баку-10,

ул. Низами, 137, бл.10, кв.90
Тел.: (8922-код) 98-30-16

Ю.А.Святец
320625 Украина, Днепропетровск, пр.Гагарина, д.72
Днепропетровский госуниверситет
исторический факультет

д.и.н. В.Н.Сидорцов
220080 Беларусь, Минск, пр-т Франциска Скорины, д.4
Белорусский государственный университет
исторический факультет

д.и.н. Т.И.Славко
620083 Россия, Екатеринбург, пр.Ленина, д.51
Уральский госуниверситет
исторический факультет

д.и.н. А.К.Соколов
Москва, Институт Российской истории РАН
Дом. адрес: 117465 Москва, ул.Теплый Стан, д.5/2, кв.107
Тел.: 337-78-86

В.И.Тихонов
109544 Россия, Москва,
ул.Международная, д.10
Мосгорархив
Центр научного использования
и публикации архивного фонда

к.и.н. В.С.Тяжелникова
109544 Россия, Москва,
ул.Международная, д.10
Мосгорархив
Центр научного использования
и публикации архивного фонда

к.и.н. И.Ф.Юшин
109544 Россия, Москва,
ул.Международная, д.10
Мосгорархив
Центр научного использования
и публикации архивного фонда

Dr. Wolfgang Levermann
Max-Planck-Institut für Geschichte
Hermann-Foge-Weg 11
D-3400 Göttingen
Germany