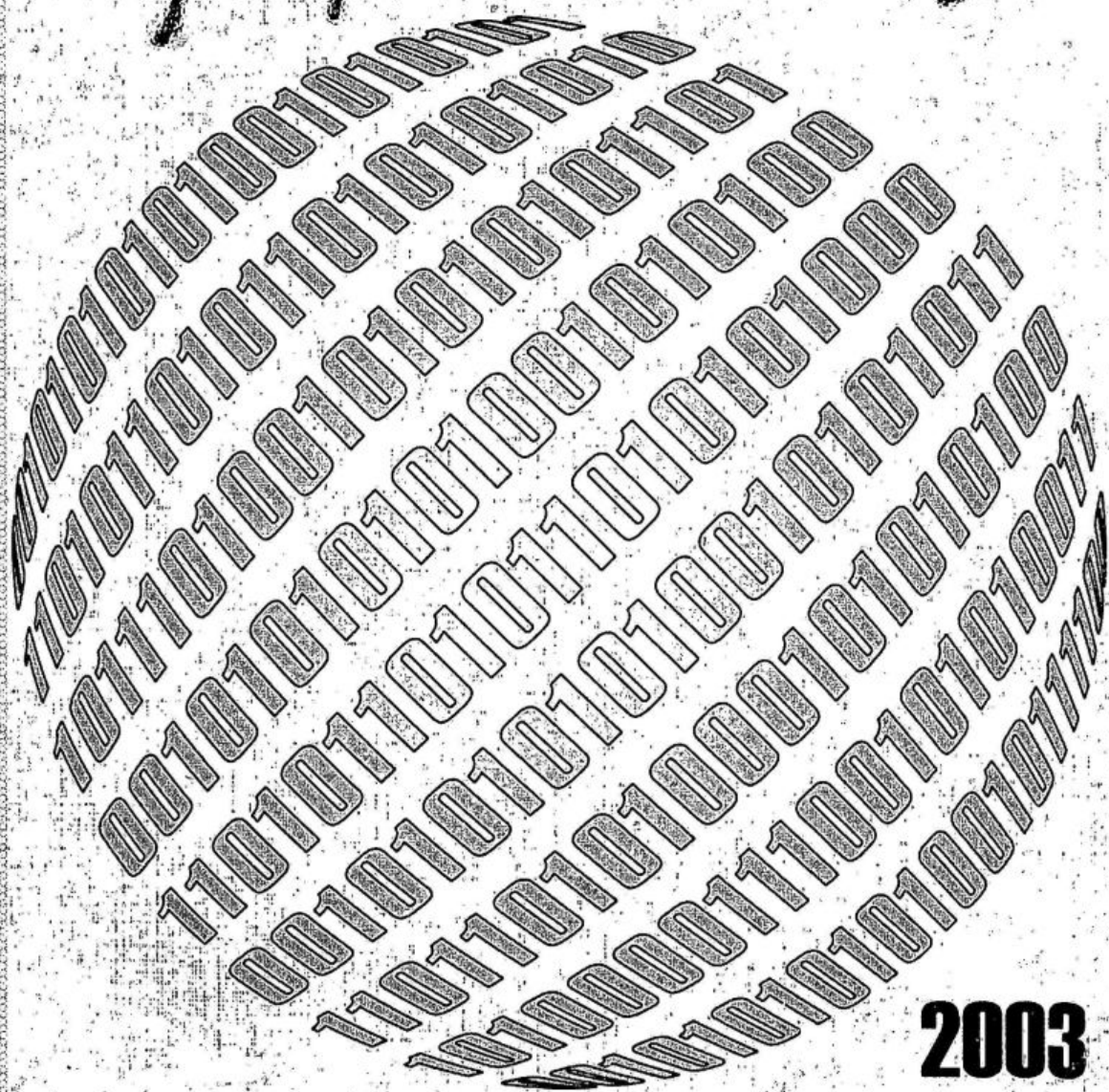




Информационные технологии в гуманитарных исследованиях



2003

РОССИЙСКАЯ АКАДЕМИЯ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ АРХЕОЛОГИИ И ЭТНОГРАФИИ

АССОЦИАЦИЯ "ИСТОРИЯ И КОМПЬЮТЕР"
СИБИРСКОЕ ОТДЕЛЕНИЕ

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ

Выпуск 6

Ответственный редактор
Академик РАН, доктор исторических наук *Ю. П. Холюшкин*

Новосибирск
2003

ББК 60
И 74

Издание осуществлено при финансовой поддержке Российского фонда фундаментальных исследований, Российского гуманитарного научного фонда и интеграционных программ Сибирского Отделения РАН

**И 74 Информационные технологии в гуманитарных исследованиях:
Сборник трудов. Выпуск 6. Новосибирск: Новосибирский
госуниверситет, 2003.100 с.**

В сборнике излагаются подходы к подготовке, созданию, обработке и представлению информации в гуманитарных науках. Сборник рассчитан на гуманитариев, математиков и на широкий круг исследователей, интересующихся информационными технологиями в гуманитарных исследованиях и образовании.

ББК 60

ISBN 5-94356-166-8

© Институт археологии и этнографии СО РАН, 2003

СОДЕРЖАНИЕ

От редактора	4
I. МЕТОДОЛОГИЯ И МЕТОДИКА АРХЕОЛОГИИ	8
Холюшкин Ю.П. Место археологической историографии в системной классификации археологической науки	8
Холюшкин Ю.П. О месте типологической археологии в системной классификации археологии	14
Васильев С.А. Статистическое изучение динамики развития древней культуры Верхнего Енисея	18
II. СТАТИСТИЧЕСКИЕ МЕТОДЫ И ГИС-ТЕХНОЛОГИИ В АРХЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ	21
Деревянко А.П., Холюшкин Ю.П., Бердников Е.В., Воронин В.Т. ГИС «Палеолит Северной Азии»	21
Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Ростовцев П.С., Костин В.С., Корнюхин Ю.Г. Статистические методы в изучении каменных индустрий	30
Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Костин В.С. Предварительные данные по структурному анализу технологических индексов мустьерских комплексов Средней Азии и Казахстана	46
Костин В.С. Статистика для сравнения классификаций	57
Костин В.С., Корнюхин Ю.Г. Построение обобщенной классификации	65
III. ИНФОРМАЦИОННЫЕ РЕСУРСЫ В ГУМАНИТАРНЫХ ИССЛЕДОВАНИЯХ	73
Холюшкин Ю.П., Гемуев И.Н., Бауло А.В., Воронин В.Т., Нуртдинов А.Н. Религиозно-мифологические представления народов Западной Сибири	73
Холюшкин Ю.П., Гемуев И.Н., Воронин В.Т., Фурсова Е.Ф., Бауло А.В., Воробьев В. В., Грищенко А. А. Межэтнические взаимодействия и межконфессиональное согласие в Сибири (электронная библиотека как инструмент решения проблемы)	77
Воронин В.Т., Холюшкин Ю.П., Бердников Е.В., Федоров С.А., Жилицкая Г.Ю. Электронный каталог научной библиотеки института археологии и этнографии СО РАН (археология и этнография)	81
IV. ПРОБЛЕМЫ ПАЛЕОАСТРОНОМИИ	86
В.Е. Ларичев. Медведь в искусстве раннего этапа верхнего палеолита Сибири (семантика образа в свете календарно- астрономических записей)	86

Очередной шестой выпуск сборника является юбилейным для сектора археологической теории и информатики, поскольку 10 лет назад по инициативе академика А.П.Деревянко при секторе палеолита была создана неструктурная группа информатики.

За время существования научного подразделения в секторе велись работы по трем направлениям:

1. Разработка системной классификации археологической науки.
2. Разработка оригинальных алгоритмов, вычислительных процедур и схем, обеспечивающих многоэтапный анализ археологических данных по разным алгоритмам и правилам.
3. Создание информационных ресурсов.

По первому направлению в секторе велись разработки системной классификации археологии и была опубликована монография Ю.П.Холюшкина и Е.Д.Гражданникова «Системная классификация археологической науки (элементарное введение в археологическое науковедение)» – Новосибирск, 2000. 60 с. - 7,3 п.л. и ряд статей.

В рамках второго направления был разработан ряд оригинальных алгоритмов, вычислительных процедур и схем, которые обеспечили многоэтапный анализ археологических данных по разным алгоритмам и правилам. Описание и теоретическое обоснование этих методов были опубликованы в ряде сборников научных трудов сотрудников сектора информатики и в коллективных монографиях, подготовленных сотрудниками САТИ.

В основе созданных алгоритмов анализа лежит комплекс взаимосвязанных идей: выявление структуры таблицы с помощью упорядочения строк и столбцов, нейтрализации неопределенных элементов таблицы, разбиение ее на связанные области, построение и применение адекватного критерия качества разбиения, выбор последовательности процедур и критерия остановки разбиения, алгоритмы перемещения границ, анализ матриц сопряженности, понятие и смысл автоматизации типологического группирования, построение критерия качества кластеризации для одномерного и многомерного случаев, для неколичественной переменной, устранение влияния малых групп, логика группирования, интерпретация результатов группирования, устойчивость выявленных структур, устойчивость в анализе структуры таблиц сопряженности, устойчивость при типологическом группировании и т.д. Описанный аналитический базис методологии статистического исследования объектов дополняется схемой применения в анализе данных метода повторной выборки с возвращением (bootstrap). Эти положения, теоретически обоснованные и сформулированные на языке алгоритмов и процедур, легли в основание новой методологии комплексного анализа, пригодной даже для «плохих» данных, какими являются данные археологических исследований.

Третьим направлением сектора является создание Электронных ресурсов сибирской и мировой археологии и этнографии". В рамках этого направления в секторе создан крупный информационный узел, включающий в себя :

1. Информационный центр Института археологии и этнографии.
2. Информационный центр «Сибирика» с рядом ассоциированных сайтов.
3. Информационный портал «Гуманитарная паутина».
4. Сайт Западно-Сибирского отделения Российской академии естественных наук.

В реализации данного проекта были использованы современные технологии создания и поддержки информационных ресурсов в сети Интернет (база данных Postgres, HTML-формы, JavaScript; язык PHP для динамического создания HTML-страничек и обработки запросов), обеспечивающих реализацию концепции динамических документов и автоматизации обновления документов. Применение этих технологий обеспечило регулярную поддержку на информационном узле раздела новостей (Института, сайта, конференций, новостей фондов, персоналий, лектория).

Была осуществлена поддержка предметно-ориентированных систем доступа к ведущим археологическим сетям, международным исследовательским центрам, библиотекам, архивам и музеям.

Кроме того, развернута работа по созданию библиографической системы археологических изданий на основе разрабатываемой в секторе системной библиографической археологической классификации. Была разработана система автоматизированной верстки электронных изданий, включающая в себя полноценные электронные публикации (журналы, бюллетени, сборники научных статей, Web-энциклопедии). Благодаря этому удалось осуществить размещение на сервере десятка справочных изданий, нескольких сотен публикаций по археологии и этнографии, ряд этнографических фильмов. Начата разработка различных информационных систем (ГИС, библиотечных и др.).

Данный сборник в какой – то степени отражает содержание работ, проводимых в последние годы в секторе археологической теории и информатике.

Первый раздел сборника посвящен проблемам, связанным с разработкой системной классификацией гуманитарных наук (на примере археологии). Предполагается использование этих разработок в создаваемом совместно с Институтом систем информатики СО РАН портале археологических знаний. В плане работ по совершенствованию системной классификации археологической науки и ее представления уже выполнено следующее:

Разработаны структуры для представления классификационных фрагментов, описывающих научную дисциплину и ее составляющие.

Разработаны и реализованы web-интерфейсы пользователя и администратора системы для возможности удаленного просмотра и ведения системной классификации. Реализация выполнена с помощью средств PHP4, Java и MySQL в качестве СУБД.

Проведено начальное построение иерархии классификационных фрагментов с помощью реализованных средств.

Второй раздел сборника посвящен статистическим методам и ГИС-технологиям в археологических исследованиях.

Первая статья раздела посвящена созданию и развитию геоинформационной системы по палеолиту Северной Азии.

Авторами проекта в течение 1990-2000 гг. проведены подобные расчеты по обработке данных материалов публикаций в научной печати по археологии палеолита Евразии, в том числе по археологии палеолита Северной Азии.

В качестве моделей использовались:

- методы и модели информационно-статистического анализа и группирования (Фелингер, 1995), построенные на теоретико-информационных основаниях (количество информации, энтропия, условная информация и энтропия, мера структурированности данных, мера неопределенности данных, мера зависимости одних признаков совокупности от других и т. д.);

- модели анализа таблиц сопряженности, типологического группирования (кластерный анализ) и анализа устойчивости выделенной структуры данных (Ростовцев и др., 1994-2000); алгоритмы моделей основаны на логике последовательного разбиения совокупности объектов на группы с последующим объединением групп в типы; выводы о статистической устойчивости результатов расчетов по моделям анализа таблиц сопряженности строятся на основе метода повторной выборки с возвращением (метод Boot Strap).

Эти исследовательские инструменты дополняются методами восполнения недостающих, пропущенных или отсутствующих данных, факторного анализа и предметной (археологической) интерпретации результатов статистической обработки.

На основе интеграции инструментария авторами проекта разработана технология глубокой эшелонированной обработки информации, результаты которой могут быть размещены и представлены в наиболее наглядной форме в археологических ГИС.

При разработке программно-технических решений по созданию ГИС авторами проекта учитывались следующие из требования:

открытый доступ по WWW к накопленным материалам и результатам исследований в области палеолита Северной Азии;

программно-информационное обеспечение ГИС на основе современных информационных технологий;

обеспечение открытости и оперативности при внесении новых данных и обновления существующих информационных ресурсов ГИС.

При разработке цифровой карты памятников палеолита Северной Азии применялись технологии, представленные в открытых публикациях. При этом были использованы современные программные средства и утилиты, в том числе настольная геоинформационная система ArcView, а также утилиты собственной разработки. Для этой цели применялся метод растровой векторизации бумажного картографического материала.

При подобной векторизации данные (которые были сформированные в процессе работы над проектом реляционной базы данных палеолитических памятников, содержащей координаты «широта-долгота») были конвертированы в рабочий векторный формат Shapefile.

В качестве топографической основы были использованы свободно-доступные векторные карты мира и Российской Федерации, а также глобальная растровая географическая карта. Полученный в результате компоновки тематических и общегеографических слоев набор данных позволяет достаточно наглядно отобразить различные атрибутивные характеристики памятников палеолита Северной Азии, их географическое положение, периодизацию и общие показатели, характеризующие пространственно-временную распределенность археологического материала.

Ряд статей этого раздела посвящено статистическому анализу палеолитических объектов. При проведении автоматической классификации часто возникает вопрос о том, насколько выделенные программой классы отражают реальную структуру данных. Исследователю необходимо получить подтверждение того, что обнаруженная кластерная структура не является случайной флуктуацией. Одним из способов убедиться в неслучайности найденной кластерной структуры является сравнение классификаций, построенных на разных признаковых пространствах.

Среди них отмечены ряд публикаций, в которых рассматриваются проблемы, связанные с методическими проблемами применения статистического анализа в археологии. В частности в статьях В.С.Костина делаются попытки решить задачу построения сводной, обобщенной классификации по результатам автоматической классификации 64-х археологических памятников, проведенной заранее разными методами (k-средних, иерархического кластерного анализа и типологического анализа) и на разных признаковых пространствах.

Третий раздел сборника посвящен проблемам, связанным с созданием информационных ресурсов в этнографии и созданию электронных библиотек. Важным направлением этих исследований и разработок является создание геоинформационной системы по религиозно-мифологическим представлениям народов Западной Сибири на основе материалов опубликованных монографий, статей, музейных коллекций Института археологии и этнографии СО РАН. Проектом создания этой системы предусмотрена организация свободного доступа пользователей по WWW к базе данных ГИС, исполнение этих запросов системой, разработка и встраивание прикладных программ для облегчения наполнения и редактирования базы данных. Создаваемая ГИС предусматривает использование в решении широкого круга задач этнографии, связанных с вовлечением в исследовательский процесс большого объема накопленных этнографических данных по религиозно-мифологическим представлениям народов Западной Сибири.

Актуальность проекта этнографической ГИС обусловлена тем, что:

1) ГИС-системы в этнографии являются мощным комплексным информационным ресурсом, объединяющим разнородные этнографические данные в наиболее естественной для пользователя форме представления;

2) в среде исследователей, преподавателей, студентов и всех интересующихся мифологическими и религиозными представлениями народов Западной Сибири быстро возрастают потребности в интерактивных комплексных формах представления данных (Интернет, сетевых ГИС, гипертекстовых баз данных);

3) в настоящее время все крупные этнографические, археологические и музейные центры, высшие учебные заведения и культурно-просветительские учреждения имеют собственные web-узлы, с помощью которых пользователи могут получать доступ к информации, размещенной в Интернет.

Две статьи раздела посвящена проектам по созданию полнотекстовых документов в Интернет по этнографической тематике и разработке электронного каталога по археологии и этнографии Сибири по изданиям и рукописям, хранящимся в научной библиотеке Института археологии и этнографии СО РАН. Создаваемый электронный каталог представляет собой информационную систему со свободным доступом к базам данных по сети Интернет для научных сотрудников, преподавателей, студентов и учащихся.

В рамках проекта предусмотрены разработка и встраивание прикладных программ для облегчения наполнения и редактирования базы данных. Ядро электронного каталога составляют библиографические данные о материалах, содержащих результаты научных исследований сотрудников Института археологии и этнографии СО РАН, хранящихся в фондах его научной библиотеки в форме редких книг, научных монографий, сборников научных трудов, периодических изданий, материалов международных, национальных и региональных научных конференций, препринтов, авторефератов диссертаций, рукописей диссертаций, собраний текстов рукописей, собранных в археографических экспедициях, и т.п. Большая часть этих материалов представлена в единичном экземпляре и не имеет подлинников и копий за пределами библиотеки.

Завершает сборник раздел, посвященный палеоастрономии. Раздел представлен полемической статьей В.Е. Ларичева, считающего, что главной задачей искусствоведения палеолита России – избавление его от пустопорожних («философических») разговоров и выбор методики, которая обеспечила бы *доказательность семантическим оценкам образов и композиций*, а с нею, доказательностью этой, достойное науки решение проблемы происхождения художественного творчества. Такой, пифагорейского стиля метод («Число объясняет Все»), был разработан в ходе изучения предметов искусства Мальты, Малой Сии, Ачинского поселения, Сунгири, Игнatieвской пещеры и bâtons de commandement, а также прочих объектов искусства западноевропейских памятников древнекаменного века, в том числе – пещерных (см., для примера [Ларичев, 1989; 1993; 2003]). Что касается образа медведя в искусстве палеолита Европы, то этому сюжету была посвящена специальная публикация (см. рис. 1, 3, к [Ларичев, 2002]; там же – обоснование неприемлемости старых методических подходов при строгой нацеленности искусствоведа на интерпретации). Это и определило выбор темы публикации – семантика образа медведя в искусстве палеолита Сибири. Она, тема эта, призвана, по его мнению, помимо прочего, наглядно продемонстрировать – какова была реальная значимость всего лишь одного «неудобного» образца искусства Малой Сии и во что обходятся делу погромы, учиняемые в ученой среде амбициозными личностями неустойчивой нравственности.

Ю.Холюшкин

Холюшкин Ю.П.

Место археологической историографии в системной классификации археологической науки

Настоящая статья является продолжением исследований по построению системной классификации археологической науки, частично реализованной в ряде наших работ [Гражданников, Холюшкин, 1990].

В ходе этих исследований стало ясно, что парадигма теоретической археологии только формируется. Несомненно, ее становление и последующее развитие может стать реальным только в недрах информатики.

Попытаемся внести ясность в понимание структуры общей теоретической археологии, приведя соответствующий классификационный фрагмент (Рис. 1). Фрагмент был создан Е.Д.Гражданниковым [1987:78] и нами были внесены изменения лишь в диадную группу.

Археологическое науковедение				
Историография археологической науки			Археологическая критикология	
Археологическая наукометрия		Археологическая критериология		Археологическое прогностическое науковедение
Археологическая информатика				
Археологическая синектика	Публикационное науковедение	Науковедение открытий	Программно- целевое науковедение	Теория развития археологического знания

Рис. 1. Классификационный фрагмент "Общая теоретическая археология"
[по Е.Д.Гражданникову, Ю.П.Холюшкину]

Здесь в качестве опорного понятия приведено археологическое науковедение. Науковедение представляет собой "историческое, философское, социологическое, операциональное и т.п. исследование науки, техники, медицины и т.п.". [Прайс, 1966: 30; Гражданников, 1987: 75]. Предметом изучения археологического науковедения как науки являются готовые научные результаты, полученные в основных ее областях, а целью - разработка методов измерения, оценки и прогнозирования научной продукции.

В качестве диадной группы приведены понятия "историография археологической науки" и археологическая критикология.

Историография археологической науки исследует историю науки и представляет собой совокупность исследований, посвященных определенной теме или исторической эпохе. Прослеживая пути исторического познания прошлого, историографические исследования позволяют выявить движущие силы прогресса науки, закономерности появления и смены археологических направлений и школ, меру объективной истинности созданных ими концепций. В качестве задач археологической историографии можно назвать следующие:

1. *Выяснение теоретико-методических принципов, присущих каждому направлению археологической мысли;*
2. *Анализ источниковедческой базы, характера использования археологических источников, конкретных методик исследования археологических фактов;*

3. *Определение места археологических научных школ и направлений в установлении и систематизации археологических фактов;*
4. *Изучение организации и форм исследовательской работы в области археологии, включая систему научных учреждений, вузов, музеев и архивов, систему подготовки кадров;*
5. *Анализ проблематики археологических исследований. Ее развития и расширения;*
6. *Исследование археологических концепций, созданных различными археологическими направлениями и школами и выявлению моментов преемственности археологического знания.*

В настоящее время стало традицией посвящать истории археологии не только статьи, но также и историографические очерки, которые почти традиционно открывают обобщающие монографии, посвященные археологическим памятникам. В этих обзорах содержатся иногда редкостные по ценности материалы, раскрывающие детали хода накопления археологических знаний в отдельных районах Северной Азии. В них можно найти также детальный перечень открытий, списки находок, «презентацию» взглядов отдельных исследователей, в том числе незаслуженно забытых. И все же полного удовлетворения от знакомства с такими, порой, увы, «дежурными» по направленности и стилю разделами публикаций не получаешь. Простая констатация известных фактов оставляет, как правило, вне внимания авторов историографических скетчей весьма интересный аспект - творческую лабораторию лидеров археологии Сибири как дореволюционной, так и современной. Между тем, как верно заметил А.А.Формозов, «едва ли не важнейшей задачей истории археологии является анализ идей и методов наших предшественников. Важно не только то, что они нашли, но и то, какие цели ставили они при своих исследованиях, как их осуществляли» [Формозов, 1975: 9]. Остаются втуне также поиски нового стиля изложения историографии, отсутствует стремление включить в них давно напрашивающиеся разделы теоретико-методического плана. Это направление можно было бы назвать содержательным науковедением, поскольку главная идея археологической историографии - дать представление об археологической науке, как результате человеческой деятельности.

Второе диадное понятие "археологическая критика" занимается критической проверкой ранее использованных приемов интерпретации археологических фактов, установления степени их достоверности и, вместе с тем, нового осмысления уже накопленных археологических материалов. Это объективно необходимо в результате обнаруживавшегося нового значения уже известных фактов, являющихся звеньями, непрерывно растущей цепи взаимосвязанных событий в истории развития науки. Довольно часто это направление использовалось не для реализации указанных выше целей, а для борьбы с инакомыслием. Между тем каждый из этих разделов выявляет определенную сторону проблемы движущих сил развития науки, взаимно дополняя друг друга.

Триадная группа была построена Е.Д.Гражданниковым. Здесь в качестве первого понятия приведена археологическая наукометрия [с2]. Наукометрия была основана Д. Дж. Прайсом в 1951 году. В качестве раздела археологической науки ее целью является статистическая обработка эмпирических археологических науковедческих данных.

Археологическая критериология - это наука о критериальной оценке научных результатов, т.е. об оценке научных результатов на основе их сравнения посредством строго определенной статистической процедуры. Попытки критериальной оценки научных результатов предпринимались нами для проверки правомерности выделения ряда археологических культур Сибири [Холюшкин, Холюшкина, 1985].

Археологическое прогностическое науковедение - раздел археологической науки о прогнозировании научно-технического прогресса, т.е. получении информации о научных результатах, которые могут быть получены в будущем

В качестве альтернативно-тождественного понятия в классификационном фрагменте выступает археологическая информатика. Под этой дисциплиной понимается наука, изучающая структуру и общие свойства научной информации, а также закономерности ее создания, преобразования, передачи и использования в различных сферах человеческой деятельности [БСЭ, Т. 10: 348].

Как писал в свое время Е.Д.Гражданников, информатике можно поставить в соответствие пятиэлементную группу: синектика (экспертное науковедение) - публикационное науковедение - науковедение открытий - программно-целевое науковедение - теорию развития науки [Гражданников, 1987: 77].

Экспертное науковедение основано на использовании экспертных оценок. Эти оценки имеют малую точность, благодаря наличию субъективных ошибок, связанных с экспертом. Здесь часто искажаются оценки работ вследствие личных причин, личных пристрастий, местного патриотизма и т.п. [Хайтун, 1983:89].

Публикационное науковедение основано на использовании статистического анализа потоков публикаций и цитирования для оценки перспективности научных направлений и определения ценности научных работ [Гражданников, 1987: 78]. Данное направление в археологии стало развиваться сравнительно недавно, но и эти отдельные разработки открыли новые возможности для проверки качества научных результатов в археологии [Деревянко, Холмошский, 1994б].

Науковедение открытий - это анализ взаимосвязанных публикационных комплексов с целью выявления результатов высокой ценности [Гражданников, 1987: 78].

Программно-целевое науковедение - разработка методов решения крупномасштабных научных задач, основанное на научном прогнозировании [Гражданников, 1987: 78].

Теория развития археологической науки основана на анализе развития мировой науки с учетом социальных последствий практического применения ее высших достижений [Гражданников, 1987: 78]. Начало развитию этому направлению положила книга Т.Куна "Структура научных революций" [1975]. Именно Т.Кун ввел в науковедение понятие парадигмы. "Научная революция" и "парадигма" - два основных понятия теории развития науки. Для раскрытия этого научного понятия приведем классификационный фрагмент "Теория развития археологической науки" (рис. 2).

Под общей теорией развития археологической науки понимается раздел теории развития археологического знания, целью которого является обобщение опыта использования методов анализа развития мировой археологической науки с учетом социальных последствий практического применения ее высших достижений.

Общая теория развития археологической науки					
Интерналистское направление археологической науки			Экстерналистское направление археологической науки		
Допарадигмальная археология		Парадигмальная археология		Постпарадигмальная археология	
Теория процесса производства археологического знания					
Классификационная стадия	Стадия систематизации	Теоретическая стадия	Догматическая стадия	Стадия замещения	

Рис. 2. Классификационный фрагмент "Теория развития археологического знания"

В диадной группе приведены два направления в историографии науки - интерналистское и экстерналистское. Каждое из них правомочно в своей сфере и выявляет определенные стороны проблемы движущих сил развития науки [Toulmin, 1972].

Интерналистское направление сосредоточено на исследовании внутридисциплинарных факторов развития науки: на выявлении линий интеллектуальной преемственности, критериев выбора приемлемой теории и т.д. Первое из этих направлений уделяет наибольшее внимание нормативной традиции, управляющей поведением ученых. Примером такого подхода является попытка

выявления преемственности при разработке схемы развития идей в палеолитоведении Северной Азии [Деревянко, Фелингер, Холушкин, 1989:239-245].

Экстерналистское направление дополняет картину развития научной дисциплины выявлением каналов ее взаимодействия с различными социальными институтами. Это направление скорее интересуется социально-психологическими процессами, посредством которых устанавливаются определенные авторитарные традиции научной мысли, становятся привычными, а затем подрываются и заменяются иными [King, 1980: 104].

Допарадигмальная археология - наука периода частых и серьезных споров о правомерности применяемых методов исследований, проблем и стандартных решений. Почти несущественные в период нормального развития науки, они регулярно вспыхивают в предпарадигмальный период. Именно в этот период ученые-археологи спорят о том, решены ли фундаментальные проблемы в их области, а поиски научных правил приобретают такое значение, которое они обычно не имели.

Парадигмальная археология - археология периода, когда научное сообщество получает по крайней мере критерий выбора проблем, которые могут считаться разрешимыми, пока эта парадигма принимается без доказательства. До тех пор, пока средства, представляемые действующей парадигмой, позволяют успешно решать проблемы, порождаемые ею, наука продвигается наиболее успешно и проникает на самый глубокий уровень явлений, уверенно используя эти средства [Кун. 1975: 105]

Постпарадигмальная археология - археология периода кризиса нормального развития науки. Значения кризисов заключаются в том, что они говорят о своевременности перестройки научной парадигмы. Любой кризис начинается с сомнения в парадигме и последующего расшатывания нормального исследования. В этом отношении исследование во время кризиса имеет очень много сходного с исследованиями в допарадигмальный период. В период кризисов многие ученые начинают постепенно терять доверие к прежним теориям и затем начинают задумываться об альтернативах для выхода из кризиса, тем не менее они никогда не отказываются легко от парадигмы, которая их ввергла в кризис.

В качестве альтернативно - тождественного понятия нами введена теория процесса производства археологического знания, которая обнаруживает связь с пентадной группой.

Классификационная стадия - период развития науки, когда ученые главным образом сосредотачивают внимание на первичных операциях группировки, сортировки археологических материалов. В России первые шаги по классификации материалов делались А.С.Уваровым и Д.Я. Самоквасовым [Формозов, 1995: 24]. В этот период возникает тенденция рассматривать археологическую науку как единое целое, соединяющее первобытные, средневековые и славяно-русские древности [Генинг, 1982]. Именно тогда создаются первые хронологические классификации археологических материалов. Однако для работ этой стадии характерно отсутствие каких либо глубоких идей. Главными для исследователей являются лишь факты и убеждение в том, что они сами все скажут за себя, сами дадут некий синтез.

Стадия систематизации - период учета основных взаимосвязей в конкретной области знания и стремления к созданию универсальных классификационных систем артефактов, соотношений объектов с классами и классов между собой. В археологии - это период культа естествознания, стремления создания систем, заимствованных во многом из биологии. В России первыми систематизаторами принято считать В.А.Городцова, А.А.Спицына и В.Сизова. Именно благодаря трудам этих исследователей было получено первое представление о предмете археологии и объекте ее исследования. Так для В.А.Городцова, археология - это "реальная, преимущественно систематическая наука о творениях и отношениях к окружающей природе вымерших поколений человечества" [Городцов, 1923: 5]. В этот же период формируется ряд

законов существования археологических памятников: индустриальной причинности, индустриальной эволюции, индустриальных заимствований, индустриальных случайных совпадений, индустриальной борьбы вещественных археологических творений за свое существование [Городцов, 1927: 4].

Теоретическая стадия - период возникновения, формирования и утверждения различных идей, концепций и научных школ. В археологии - это диффузионизм, школа культурных кругов, культурно-историческая школа, палеоэтнологическое направление и, наконец, социологическое направление, нашедшее свое воплощение в теории стадиальности.

Догматическая стадия развития науки - это период развития науки, когда ученые концентрируют свое внимание на проблемах, решению которых может помешать только недостаток собственной изобретательности. В этот период научное сообщество часто самоизолируется от насущных социальных и научных проблем, которые нельзя свести и представить в терминах концептуального и инструментального аппарата, предлагаемого действующей парадигмой [Кун, 1975: 60]. Такие проблемы рассматриваются лишь как отвлекающие внимание исследователей от подлинных научных проблем. В России эта стадия была названа в одной из статей Л.С.Клейна периодом талмудизма.

Стадию замещения старой парадигмы можно назвать революцией. В этот период изменяются не только проблемы, но и вся сеть фактов и теорий претерпевают такую замену. И они являются основой для новой традиции нормальной науки. Как указывал Т.Кун, здесь возможно несколько исходов, включая сохранение старой парадигмы. Но чаще всего происходит появление нового претендента на роль новой парадигмы и последующей борьбой за ее принятие [Кун, 1975: 115]. В этом пункте сообщество ученых разделяется на враждующие лагеря. При такой поляризации выбор между конкурирующими парадигмами подобен выбору между политическими институтами. В этой борьбе каждая группа использует свою собственную парадигму для аргументации в защиту этой же парадигмы, включая и силу. Примером такой борьбы было насильственно насаждавшееся «новое учение о языке» Н.Я. Марра, навязывание ученым принципов автохтонизма и стадиальности, уничтожение в России палеоэтнологического направления.

Описанное является крайностью в истории науки, чаще этот процесс, по Т.Куну, скорее напоминает реконструкцию научной области на новых основаниях, которая изменяет некоторые наиболее элементарные теоретические обобщения в данной области, а также многие методы и приложения старой парадигмы. Но и в крайнем случае новая парадигма сохраняет значительное число старых достижений. Так, в предвоенные годы была выдвинута на важные позиции идея преемственности науки и хотя, по-прежнему, важнейшей целью археологии признается изучение социально-экономической проблематики прошлого, уже не отрицалась публикация отдельных археологических материалов и классификаций древних изделий [Формозов, 1995: 59].

Приведенная классификация может создать впечатление о том, что мы пытаемся представить историю науки в линейном и кумулятивном виде. В действительности она позволяет говорить о перекрывании периодов, влиянии на перестройки в науке внутренних и внешних (социальных, психологических, политических и др.) факторов на всех стадиях развития конкретной науки, о чем говорит и положение понятий в классификационном фрагменте.

Рассмотрим классификационный фрагмент «Историография археологии» (Рис. 3).

В классификационном фрагменте в качестве опорного понятия приведена «Общая историография», задачей которой является обобщение путей исторического познания прошлого, движущих сил прогресса археологической науки, закономерностей появления и смены археологических направлений и школ, мер объективной истинности созданных ими концепций.

Общая историография археологической науки					
Комментарийная историография			Теоретическая историография		
Хроникальная историография		Источниковая историография		Концептуальная историография	
Конференции	Персоналии	Археологические публикации	Археологические факты	Научные учреждения	Научные школы
Методика историографии					
Локальная историография	Региональная историография	Страноведческая историография	Континентальная историография	Мировая историография	

Рис. 3. Классификационный фрагмент "Историография археологии"

В качестве первого диадного понятия приведена «Комментарийная археологическая историография», представляющая наиболее распространенный вид историографического описания основанного на простой констатации известных фактов. При этом все рассуждения, пояснительные и критические замечания о чём-нибудь считаются излишними, т.к., по мнению этих исследователей, всё понятно и объяснения не нужны.

В качестве второго диадного понятия приведена «теоретическая историография», задачей которой является анализ источниковедческой базы, характера использования археологических источников, конкретных методик исследования археологических фактов;

определение места археологических научных школ и направлений в установлении и систематизации археологических фактов;

изучение организации и форм исследовательской работы в области археологии, включая систему научных учреждений, вузов, музеев и архивов, систему подготовки кадров;

анализ проблематики археологических исследований, ее развития и расширения.

Хроникальная историография - представляет собой раздел историографии, занимающийся описанием археологических исследований в их хронологическом порядке. Почти каждое монографическое или диссертационное исследование начинается с описания истории исследования памятника, географического района в хронологическом порядке.

Источниковая историография – отрасль археологической историографии, изучающая историю археологического знания на основе вещественных и письменных источников.

Концептуальная историография - отрасль археологической историографии, изучающая теоретико-методические принципы, присущие каждому направлению археологической мысли и исследованием археологических концепций, созданных различными археологическими направлениями и школами, выявлением моментов преемственности археологического знания.

В качестве диадно-триадных понятий в классификационном фрагменте приведены: историографические следующие подразделы:

конференции - отчеты о проведенных конференциях;

персоналии - юбилейные статьи, некрологи, статьи об исследователях теоретического характера, в которых осуществляется анализ идей и методов наших предшественников;

публикации - обзоры археологических публикаций;

археологические факты - анализ археологических фактов;

пятое понятие «научные учреждения» отражает изучение типов организации и форм исследовательской работы в области археологии, включая систему научных учреждений, вузов, музеев и архивов, систему подготовки кадров;

шестое понятие «научные школы» представляет направление, занятое исследованием археологических научных школ в процессе их возникновения, расцвета и упадка, а также вопросов преемственности, различных форм взаимодействия.

Обратимся к методике историографии. Под археологической историографической методикой понимается последовательность конкретных процедур, реализация которых обеспечивает достижение заданной цели историографического исследования.

Пятиэлементной классификационной группой являются пять разделов археологической историографии, различающиеся масштабами объекта исследования,

начиная с локальной и кончая мировой археологией. Первая имеет дело с отдельными памятниками, а последняя с обобщением историографического мирового опыта.

Литература

- Генинг В.Ф. Очерки по истории Советской археологии. - Киев, 1982.
- Городцов В.А. Археология. Т. I Каменный период. М.-Л., 1923.
- Городцов В.А. Типологический метод в археологии. - Рязань, 1927.
- Гражданников Е.Д. Метод построения системной классификации наук. - Новосибирск, 1987.
- Гражданников Е.Д., Фелингер А.Ф., Холюшкин Ю.П. Проблемы системной классификации разделов археологии. // Проблемы антропологии и этнографии. - Иркутск, 1987: 16 - 18.
- Гражданников Е.Д., Фелингер А.Ф., Холюшкин Ю.П. Системная классификация разделов археологии. // Методические проблемы реконструкций в археологии и палеоэкологии. - Новосибирск, 1989: 5 - 16.
- Гражданников Е.Д., Холюшкин Ю.П. Системная классификация социологических и археологических понятий. - Новосибирск, 1990.
- Гражданников Е.Д., Холюшкин Ю.П. Системное решение вопроса о статусе археологии как науки [к дискуссии о предмете археологии]. // Советская археология. - М. 1991, № 2: 111 - 114.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Методы информатики в археологии каменного века. - Новосибирск, 1989.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Социологический анализ кадров археологов Сибири. // Методические проблемы реконструкций в археологии и палеоэкологии. - Новосибирск, 1989: 33-43.
- Деревянко А.П., Холюшкин Ю.П. Некоторые подходы к изучению закономерностей развития археологии Северной Азии. // Методология и методика археологических реконструкций. - Новосибирск, 1994а.
- Деревянко А.П., Холюшкин Ю.П. Проблема качественного анализа археологических публикаций. // Методология и методика археологических реконструкций. - Новосибирск, 1994б.
- Клейн Л.С. Археологические источники. - Л., 1995 [2 изд.].
- Кун Т. Структура научных революций. - М., 1975
- Прайс Д. Наука о науке // Проблемы науковедения [наука о науке]. - М., 1966: 30.
- Формозов А.А. Некоторые итоги и задачи исследования в области истории археологии // СА., №4, 1975.
- Формозов А.А. Русские археологи до и после революции. - М., 1995.
- Хайтун С.Д. Наукометрия. - М., 1983.
- Холюшкин Ю.П., Холюшкина В.А. Методологические аспекты исследования археологических культур каменного века Сибири. // Проблемы реконструкций в археологии. - Новосибирск, 1985: 23 - 45.
- King M.D. Reason, tradition and progressiveness of science.// Paradigms and revolutions. Notre Dame; London, 1980: 196-207.
- Toulmin S. Human understanding. -Oxford, 1972, vol. 1.

Холюшкин Ю.П. О месте типологической археологии в системной классификации археологии

"Основой всякой классификации является сравнение объектов, в результате которого можно сказать, что объекты либо сходны, либо различны. Понятие *сходство* и *различие* являются основными в этой процедуре" [Каменецкий, Маршак, Шер 1975: 47-48]. При этом термин «классификация» содержит в себе возможность неоднозначной трактовки. Его можно понимать и как процесс построения и как процесс использования уже построенной классификации.

В приведенном ниже, классификационном фрагменте раскрывается и расширяется понятие классификационной археологии (Рис.1).

Общая классификационная археология				
Монотетическая классификация		Политетическая классификация		
Дескриптивная классиф. археология	Сравнительная классиф. археология	Функциональная классиф. Археология		
Типологическая археология				
Аналитическая классиф. археология	Партитивная классиф. археология	Морфологическая классиф. археология	Предметная классиф. археология	Генеральная классиф. Археология

Рис. 1. Классификационный фрагмент "Теоретическая археология" (А.П.Дервянко, Ю.П.Холушкин, В.Т.Воронин, 1990).

Здесь в качестве опорного понятия приведена общая классификационная археология, в задачу которой входит разработка логических правил операций, систематики археологических объектов, их группирования, сортировки.

Диадная группа классификационной археологии построена по принципу существования регулярности в археологическом материале. Термины были введены Д.Кларком в его монографии "Analytical archaeology", согласно которой в монотетических классификациях признаки археологических объектов жестко связаны в сочетания и эти признаки резко разделены. В идеале, как полагает Д.Кларк, все признаки данного набора в каждом данном классе присутствуют. В политетической классификации группировка рассеяна и в каждом объекте присутствует лишь часть набора признаков и ни один признак не представлен у всех объектов. При подобной классификации наборы признаков у объектов различны [Clarke, 1968:3, p.35-37].

Описательная (дескриптивная) классификационная археология отражает проблематику обоснования выбора средств представления археологических объектов в терминах данного описательного языка (естественного, документального или информационного) [Гарден, 1983:52]. Как правило, подобные описательные конструкции, будучи сравнительно простыми, представляют собой последовательности высказываний, относящихся к характеристикам артефактов, и соображений о том, как эти типы могут быть использованы для объяснения археологических данных. Примерами дескриптивной классификации являются списки Борда.

Сравнительная классификационная археология связана с поисками критериев культурных сравнений с помощью сети конфигураций поведения типов и контекста их нахождения, а также с разработкой мер сходства и различия между различными типами и системами классификаций [Гарден, 1983; Клейн, 1991].

Функциональная археологическая классификация представляет собой направление, приверженцы которой занимаются разработкой методов по выяснению функциональных назначений древних вещей.

Общая классификационная археология				
Монотетическая классификация		Политетическая классификация		
Дескриптивная классификационная археология	Сравнительная классификационная археология	Функциональная классификационная археология		
Типологическая археология				
Аналитическая классификационная археология	Партитивная классификационная археология	Морфологическая классификационная археология	Предметная классификационная археология	Генеральная классификационная археология

Рис. 2. Классификационный фрагмент "Классификационная археология"

В качестве альтернативного - тождественного понятия в классификационном фрагменте выступает типологическая классификация, содержанием которой является

выявление глубинной структуры археологического материала, определение места объектов в системе, их значимости, соотношения типов с культурой.

Пентадная классификационная группа основана на представлениях Д.Кларка [Clarke, 1968], хорошо описанных Л.С.Клейном [Клейн, 1991: 545], и характеризуется поэтажной, многоуровневой организацией.

Аналитическая классификация, размещенная на первом уровне, связана с методами анализа на уровне признаков (атрибутов). Классификация следующего уровня (партитивная) предназначена для разработки методики анализа на уровне деталей. Этажом выше морфологическая классификация представляет уровень классификации специфических сторон классифицируемых предметов, таких как морфология, технология, мотивы орнамента, композиция и т.д. Еще выше уровень полного учета внутренних структур предмета (предметная классификация). И, наконец, выше всех располагается, классификация, основанная на учете всех признаков, в том числе и внешних связей (генеральная классификация, типы Кригера) [Клейн, 1991: 55].

Общая типологическая археология				
Формальная типология		Содержательная типология		
Определение	Анализ		Интерпретация	
Этапы типологизации				
Знакомство с материалом	Описание конечным набором признаков	Выбор основания типологизации	Распределение объектов в группы	Использование типологии для получения нового знания

Рис. 3. Классификационный фрагмент "Типологическая археология"

Приведем новый классификационный фрагмент «типологическая археология».

Здесь в качестве опорного понятия приведена «общая типологическая археология», задачей которой является обобщение опыта построения типологий в археологии.

В диадной группе в качестве первого диадного понятия представлена «формальная классификация», на основе которой археологи стремятся выявить узловые пункты изменчивости материала, группируя его по сходству параметров вокруг этих пунктов, не обязательно весь, и для нее важны границы между типами" [Клейн, 1987]. По своей природе формальные типологии не стремятся выявить в материале существующие в нем разграничения, а чаще всего навязывают ему несвойственные ему деления.

Здесь же в качестве второго диадного понятия представлена «содержательная классификация, в рамках которой типолог "ставит задачу открыть и охарактеризовать существовавшие в реальной исторической действительности группы." [Грязнов 1969: 19]. Задачей этого раздела является создание типологий, которые открывают реально существующие в материале типы, т.е. культурные типы.

В качестве первого триадного понятия взято «определение», задачей этой аналитической стадии типологического анализа является разбивание почти безграничного разнообразия признаков на легко различимые группы и попытки выработать отличительные признаки этих групп.

Второе триадное понятие «анализ» отражает этап выявления порядка в соотношении признаков. На этом этапе исследователь должен создать типологию.

Третье триадное понятие «интерпретация». На этом этапе типолог пытается выяснить, как и откуда произошли типы, каковы их «взаимоотношения» и выявить причины их изменений,

При этом у исследователя существует много потенциальных возможностей интерпретаций конкретных типологических групп, которые в каждом отдельном случае приходится решать отдельно.

Альтернативно-тождественным понятием является понятие «этапы типологизации». Сама процедура создания типологических классификации как определенная последовательность операций, обязательных при любом подходе к классификации, представлена в пентадной группе. Она отражена достаточно ясно в

работах разных авторов [например, Clarke 1968; Каменецкий, Маршак, Шер 1975 и др.] и выглядит следующим образом [Колпаков, 1991:72]:

1. Знакомство с материалом, включающего не только его осмотр, но и знакомство с соответствующей литературой, работу с музейными коллекциями, близкими к изучаемому материалу, и т.п. Таким образом, уже самый первый этап классификационной процедуры - выявление свойств и формирование признаков - достаточно сложен и сам включает в себя множество "пробных" классификаций. Результатом первого этапа процедуры является набор признаков и "описания" исследуемых объектов в рамках этого набора.

2. Следующий шаг - отбор признаков и их сочетаний, взвешивание признаков.

3. Основание типологии, является центральным для традиционной типологии, ибо все существенные характеристики связываются в типологии с особенностями ее основания. Поэтому хорошей типологией является та, которая объединяет в один тип артефакты, максимально сходные друг с другом в существенных признаках, является устойчивой и в тоже время достаточно гибкой, чтобы сохраняться в условиях появления все новых и новых объектов исследования.

4. Собственно типологизирование, т.е. распределение исследуемых объектов в группы по основанию классификации; (агломеративная стратегия Дорана и Ходсона В каждой конкретной классификации эти этапы наполняются конкретным содержанием. Например, агломеративная и распределительная стратегии Дорана и Ходсона.

5. Использование построенной классификации с целью получения нового знания.

Подводя итоги, отметим, что парадигма классификационной археологии только формируется. Несомненно, ее становление и последующее развитие может стать реальным только в недрах информатики.

Литература

- Генинг В.Ф. Очерки по истории Советской археологии. - Киев, 1982.
- Городцов В.А. Археология. Т.I Каменный период. М.-Л., 1923.
- Городцов В.А. Типологический метод в археологии. - Рязань, 1927.
- Гражданников Е.Д. Метод построения системной классификации наук. - Новосибирск, 1987.
- Гражданников Е.Д., Фелингер А.Ф., Холюшкин Ю.П. Проблемы системной классификации разделов археологии. // Проблемы антропологии и этнографии. - Иркутск, 1987: 16 - 18.
- Гражданников Е.Д., Фелингер А.Ф., Холюшкин Ю.П. Системная классификация разделов археологии. // Методические проблемы реконструкций в археологии и палеоэкологии. - Новосибирск, 1989: 5 - 16.
- Гражданников Е.Д., Холюшкин Ю.П. Системная классификация социологических и археологических понятий. - Новосибирск, 1990.
- Гражданников Е.Д., Холюшкин Ю.П. Системное решение вопроса о статусе археологии как науки [к дискуссии о предмете археологии]. // Советская археология. - М. 1991, № 2: 111 - 114.
- Грязнов М.П. Классификация, тип, культура // Теоретические основы советской археологии. Т.Д. - Л., 1969: 18-22.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Методы информатики в археологии каменного века. - Новосибирск, 1989.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Социологический анализ кадров археологов Сибири. // Методические проблемы реконструкций в археологии и палеоэкологии. - Новосибирск, 1989:33-43.
- Деревянко А.П., Холюшкин Ю.П. Некоторые подходы к изучению закономерностей развития археологии Северной Азии. // Методология и методика археологических реконструкций. - Новосибирск, 1994а.
- Деревянко А.П., Холюшкин Ю.П. Проблема качественного анализа археологических публикаций. // Методология и методика археологических реконструкций. - Новосибирск, 1994.б
- Каменецкий И.С., Маршак Б.И., Шер Я.А. Анализ археологических источников. - М., 1975.
- Клейн Л.С. Археологические источники. - Л., 1995 [2 изд.].
- Клейн Л.С. Классификация и типология // Исторические чтения памяти Михаила Петровича Грязнова. Т.Д, Т.1 - Омск, 1987:38-40
- Кун Т. Структура научных революций. - М., 1975

- Прайс Д. Наука о науке // Проблемы науковедения [наука о науке]. - М., 1966: 30.
- Хайтун С.Д. Наукометрия - М., 1983
- Холушкин Ю.П., Холушкина В.А. Методологические аспекты исследования археологических культур каменного века Сибири. // Проблемы реконструкций в археологии. - Новосибирск, 1985: 23-45.
- Clarke D.L. Analytical archaeology. - L.: Methuen, 1968.
- King M.D. Reason, tradition and progressiveness of science. // Paradigms and revolutions. Notre Dame; London, 1980: 196-207.
- Plog F.T. Laws, systems of law and the explanation of observed variation. // The explanation of culture change: Models in prehistory. - L., 1973: 649-661.
- Plog F. Systems theory in archaeological research // Annual Review of anthropology. - 1975, № 4: 207-224.
- Redman Ch. L. Research and Theory in Current Archaeology: An Introduction // Research and Theory in Current Archaeology. - N.Y., L., Sydney, Toronto, 1973.
- Robinson W.S. A method for chronologically ordering archaeological deposits // American Antiquity. - 1951. - Vol. 17
- Rouse I. Analytic, synthetic and comparative archaeology. // Research and Theory in Current Archaeology. - N.Y., 1973.
- Salmon M.H. Philosophy and archaeology. - N.Y., 1982 - XI: 203.
- Schiffer M.B. Archaeological Method and Theory, Volume 1, reviewed by T.G. Baugh in Journal of Field Archaeology. // Journal of Field Archaeology, 1991, V.18, №4.
- Toulmin S. Human understanding. - Oxford, 1972, vol. 1.

Васильев С.А. Статистическое изучение динамики развития древней культуры Верхнего Енисея*

Статистическое изучение динамики развития древней культуры на протяжении больших временных отрезков является одной из основных задач первобытной археологии. При этом особый интерес представляет выявление рубежей культурных изменений и прослеживание их связи с переменами в окружающей среде. Самую благоприятную возможность для подобного рода исследований представляют редкие многослойные стратифицированные памятники, снабженные серийными абсолютными датировками и суммой естественнонаучных данных, позволяющих надежно привязать археологические комплексы к временной шкале и изучить процесс взаимодействия с природной средой.

Предлагаемая работа во многом нова для мировой археологии. Несмотря на то, что исследователи палеолита постоянно имеют дело в своей практике с изучением многослойных памятников (особенно благоприятны в этом плане пещерные комплексы), задача статистического исследования большого по объему массива данных, относящихся к характеристикам каменного инвентаря, ставилась крайне редко. Это удивительно, если учесть, что только археология палеолита имеет уникальную возможность проследить динамику эволюции древнейшей культуры на протяжении огромных хронологических отрезков. Данный факт обычно рассматривается как один из привлекательных в интеллектуальном плане моментов археологии палеолита по сравнению с изучением позднейших эпох. Более того, именно в исследованиях подобного рода видится основной вклад палеолитоведения в теорию и историю мировой культуры.

Представленные в литературе исследования носят по большей части характер отдельных, не связанных друг с другом опытов, предпринимавшихся в разные годы археологами на материалах памятников различных эпох и континентов.

Вероятно, первый опыт подобного рода принадлежит Д. Коллинзу, представившему на серии графиков процентное соотношение разновидностей орудий для некоторых ориньякских и солютрейских комплексов юго-запада Франции (Collins,

* Работа выполнена при финансовой поддержке РФФИ (проект № 03-06-80418)

1965). Дальнейший шаг в разработке тематики сделал Ч. Мак-Берни, заинтересовавшийся корреляцией каменного инвентаря с преобладающими на памятнике видами фауны (McBurney, 1973). Если Д. Коллинз визуально оценивал нарастание или убывание доли тех или иных орудий, то Ч. Мак-Берни применил простой коэффициент корреляции для оценки связи параметров. Зависимости между классами орудий и промысловыми видами были проанализированы на материалах многослойной капсийской стоянки Али Таппе, восточно-оранских комплексах Хауа Фтеа и взятых из литературы параметрах Молодовы V. Была сделана также попытка вскрыть зависимость между видами орудий и разновидностями сырья (по материалам из мустьерских комплексов Котт Сен-Брелад). В итоге изучения столь прихотливо подобранных примеров выяснилось, что некоторые разновидности орудий обнаруживают положительную корреляцию с распространенностью тех или иных видов животных, в то время как другие демонстрируют устойчивую изменчивость во времени.

Наиболее детальные исследования временной вариации каменного инвентаря были проведены в 70-е гг. группой исследователей, работавших с материалами многослойной пещерной стоянки Кзар Акил в Ливане (Azougy, Hodson, 1973). Для нашей темы важно, что в данном случае в анализе были задействованы показатели, относящиеся к разным уровням анализа каменной индустрии. Так, сопоставление комплексов велось по видам ядрищ, заготовок, площадок заготовок, сокращенному и "длинному" вариантам типлистов. В качестве метода использовался так называемый "конstellационный анализ" (Doran, Hodson, 1975). Выяснилось, что для хронологического членения материала подходит даже сокращенный вариант типлиста, без дробной типологической разбивки, а временная изменчивость может быть выявлена по самым различным параметрам. Применение сходной методики к анализу эпиграветтийских комплексов Италии (Ammerman, Hodson, 1972) показало, что типологический состав инвентаря (использовался типлист Лапласа) в основном зависит от хронологической позиции памятника.

Позднее, в 80-е гг. данная тематика порой привлекала внимание в связи с интерпретационными проблемами. Я. Козловский, изучая пространственно-временную изменчивость верхнепалеолитических индустрий Европы на разных классификационных уровнях, обратился к вопросу о вычленении в массе переменных наиболее чувствительных хронологических индикаторов (Kozlowski, 1980). По его мнению, подкрепленному результатами собственного анализа серии граветтийских памятников Средней Европы, данными Й. Хана по ориньяку юга Германии, А. Дельпорта по ориньякским слоям в Ля Ферраси и др., таковыми являются редкие специфические типы, облик которых связан с социокультурным фактором.

Д. Кларк, изучая корреляцию разновидностей каменного инвентаря (как в плане общего состава инвентаря, так и по типологическим группам орудий) и фаунистических остатков из верхнепалеолитических слоев пещеры Ла Риера в Испании, пришел к существенно иным выводам (Clark, 1989). По результатам факторного анализа им было установлено наличие двух групп комплексов - с преобладанием пластинок с притупленным краем и зубчато-выемчатых форм. При этом распределение комплексов двух типов по разрезу не подчинялось временной последовательности и не зависело от культурно-исторического членения колонки. Вероятно, такое чередование отражало попеременное разнофункциональное использование пещеры.

Как мы видим из краткого и, вероятно, неполного обзора, вопросы изучения временной изменчивости каменного инвентаря далеки от своего решения. Имеющиеся данные выборочны, а выводы противоречивы. Не разработана также программа статистической обработки с учетом того, что временная последовательность сравниваемых комплексов задана изначально самой стратиграфией памятника. Обычно при сопоставлении комплексов применяют те же стандартные методы многомерной статистики (факторный, кластерный анализы, метод главных компонент, многомерное

шкалирование), что и для сопоставления комплексов в синхронном аспекте или наборов инвентаря, происходящего с различных участков памятника.

Литература

- Ammerman A. J., Hodson F. R.** 1972. Constellation analysis: a study of Late Palaeolithic assemblages in Italy // *Rivista di scienze preistoriche*, vol. 27, N 2, P. 323-344.
- Azoury I., Hodson F. R.** 1973. Comparing Palaeolithic assemblages: Ksar Akil, a case study // *World archaeology*, vol. 4, N 3, P. 292-306.
- Clark G. A.** 1989. Romancing the stones: biases, style and lithics at La Riera // *Alternative approaches to lithic analysis. Archaeological papers of the American Anthropological Association*, N 1. Tulsa: University of Tulsa, P. 27-50.
- Collins D. M.** 1965. Seriation of quantitative features in Late Pleistocene stone technology // *Nature*, vol. 205, P. 931-932.
- Doran, J. E., Hodson, F. R.** 1975. *Mathematics and computers in archaeology*. Cambridge: Cambridge University Press.
- Kozłowski J. K.** 1980. Technological and typological differentiation of lithic assemblages in the Upper Palaeolithic: an interpretation attempt // *Unconventional archaeology: new approaches and goals in Polish archaeology*. Wrocław: PAN, P. 33-55.
- McBurney C. B. M.** 1973. Measurable long term variations in some Old Stone Age sequences // *The explanation of culture change*. London: Duckworth, P. 305-315



СТАТИСТИЧЕСКИЕ МЕТОДЫ И ГИС-ТЕХНОЛОГИИ В АРХЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Деревянко А.П.
Холюшкин Ю.П. ГИС «Палеолит Северной Азии» *
Воронин В.Т.
Бердников Е.В.

В институте археологии и этнографии (ИАЭТ) СО РАН уже в течение последних десяти лет проводятся исследования и разработки в рамках комплексной программы информационного обеспечения научных исследований.

Важным направлением этой исследовательской программы является проект по созданию и развитию геоинформационной системы по археологии Северной Азии.

Геоинформационная система по палеолиту Северной Азии разрабатывается на основе материалов опубликованных данных и их статистической обработки. Проект направлен на организацию свободного доступа пользователей по WWW к базе данных ГИС, исполнение этих запросов системой, разработка и встраивание прикладных программ для облегчения наполнения и редактирования базы данных. Создаваемая ГИС предусматривает использование в решении широкого круга задач археологии палеолита Северной Азии, связанных с вовлечением в исследовательский процесс большого объема накопленных археологических данных в различных формах представления: материалы по истории исследования, фотографии и рисунки находок, топографические планы, стратиграфия. Система ориентирована на стандартные приемы и средства разработки ГИС, представление в Интернет, на использование источников данных в виде опубликованных материалов археологических исследований и наполнение данными результатов их математико-статистической обработки.

Известные различные по тематике исследований зарубежные археологические проекты, использующие технологические средства ГИС: Швеция (исторические исследования), США (ландшафтная археология Олдувайского ущелья), ориентированы в большей мере непосредственно на описание археологических памятников и археологических находок или форму их представления на территории (количество памятников, плотность их размещения и т.д.). В то же время созданы и применяются в других областях (геология, география) и в образовательных целях более развитые сетевые ГИС. Имеется отчетливая тенденция привлечения новых материалов для размещения в ГИС, в том числе материалов обработки опубликованных данных. По этому критерию разработка ГИС "Палеолит Северной Азии" соответствует мировому уровню.

Количественный анализ (в том числе математико-статистическая обработка) как методологическая основа формализации технологии получения археологического знания направлен на то, чтобы придать сущности, формам, свойствам и характеристикам археологических объектов внешнюю определенность, построить базу для количественного соизмерения и сопоставления их признаков.

Для комплексных многопараметрических задач одним из наиболее важных свойств хорошей модели является наглядность представления информации (входной, промежуточной и итоговой) для пользователя, в первую очередь для предметных специалистов (археологов), преподавателей, учащихся и интересующихся археологической тематикой. Здесь трудно найти более наглядную форму представления данных подобной информации по сравнению с ГИС, поскольку археологические данные отражают в первую очередь пространственно-временную

* Работа выполнена при финансовой поддержке РФФИ (проект № 01-07-90393).

распределенность материала исследований. Для разнородных данных, какими являются данные археологических исследований, наиболее часто используются модели информационно-статистических расчетов.

Авторами проекта в течение 1990-2000 гг. проведены подобные расчеты по обработке данных материалов публикаций в научной печати по археологии палеолита Евразии, в том числе по археологии палеолита Северной Азии.

В качестве моделей использовались:

- методы и модели информационно-статистического анализа и группирования (Фелингер, 1995), построенные на теоретико-информационных основаниях (количество информации, энтропия, условная информация и энтропия, мера структурированности данных, мера неопределенности данных, мера зависимости одних признаков совокупности от других и т. д.);

- модели анализа таблиц сопряженности, типологического группирования (кластерный анализ) и анализа устойчивости выделенной структуры данных (Ростовцев и др., 1994-2000); алгоритмы моделей основаны на логике последовательного разбиения совокупности объектов на группы с последующим объединением групп в типы; выводы о статистической устойчивости результатов расчетов по моделям анализа таблиц сопряженности строятся на основе метода повторной выборки с возвращением (метод Boot Strap).

Эти исследовательские инструменты дополняются методами восполнения недостающих, пропущенных или отсутствующих данных, факторного анализа и предметной (археологической) интерпретации результатов статистической обработки.

На основе интеграции инструментария авторами проекта разработана технология глубокой эшелонированной обработки информации, результаты которой могут быть размещены и представлены в наиболее наглядной форме в археологических ГИС.

В рамках проекта на основе применения этих инструментов в расчетах были выделены общие показатели, характеризующие пространственно-временную распределенность археологического материала (информационно-статистические группы, связанные области, группы, типы, а также показатели устойчивости разбиения памятников и артефактов на эти связанные области, группы, типы).

Основные усилия разработчиков в соответствии с планом исследований и разработок по программе проекта были направлены на решение трех комплексных задач:

- 1) доработка существующих и модернизация созданных прикладных программ, реализующих функции ГИС, отработка пилотного варианта ГИС;
- 2) пополнение базы данных по археологическим памятникам Северной Азии;
- 3) проведения опытной эксплуатации пилотного варианта ГИС; анализ результатов испытаний пилотного варианта ГИС и доработка системы с учетом выявленных отклонений от проектных решений.

При этом на начальном этапе разработки ГИС главное внимание участников проекта было направлено на решение первой задачи, обеспечивающее технологическое (программно-техническое) обеспечение проекта в целом.

В рамках этой задачи выполнены модернизация программного обеспечения, разработка программного инструментария.

Была модифицирована структура базы данных с целью применения для разработки ГИС-системы технологий, основанных на XML. Были разработаны DTD (Document Type Definition) палеолитического памятника. Для редактирования XML-файлов базы данных была разработана модель интерфейса редактора и на ее основе создана утилита на языке Java, позволяющая редактировать атрибутивные данные о памятниках палеолита. Реляционная база данных была перенесена на платформу СУБД Oracle. Для доступа к ней пользователей Интернет, были созданы скрипты на языке JSP и HTML-формы, реализующие доступ пользователей к базе. Также были разработаны стили XSL для отображения в формате HTML результатов запросов к реляционной базе данных и базе данных XML.

Основываясь на опыте создания и эксплуатации пилотного веб-сайта проекта, был модифицирован и расширен веб-интерфейс гипертекстовых страниц сайта. Для автоматизации процесса разработки веб-сайта и публикации цифровых карт в Интернет, были разработаны шаблоны веб-страниц и шаблоны для наборов пространственных данных сетевой ГИС.

Важную роль в качестве одной из методологий отображения и анализа палеолитических культур в их географическом и временном взаиморасположении авторами была отведена ГИС TimeMap. Кроме стандартных функций компоновки и просмотра цифровых карт, с помощью TimeMap исследователю предоставляется возможность создания цифровых карт с привязкой географических объектов к временным периодам, соответствующим их датировке. В качестве примера объектов, допускающих такую пространственно-временную привязку, археологические памятники наряду с памятниками архитектуры или масштабными инженерными сооружениями являются наиболее адекватными пространственно локализованными объектами, обязанными своим возникновением человеку. С помощью временных фильтров ГИС TimeMap позволяет отображать на цифровой карте объекты, относящиеся к определенному историческому периоду, представляя их в наглядной и удобной для исследования форме. В рамках пробной эксплуатации TimeMap были проведены следующие работы. Цифровые карты памятников палеолита были экспортированы в систему TimeMap. На основе базовых картографических слоев, свободно доступных в рамках международного проекта Electronic Cultural Atlas Initiative (ECAI) и тематических слоев, созданных в рамках проекта "ГИС Палеолит Северной Азии", была создана цифровая карта памятников палеолита, отражающая их периодизацию. Для этого дополнительно была использована база атрибутивных данных с датировками культурных горизонтов. Кроме того, созданный набор данных зарегистрирован в проекте ECAI.

В рамках второй задачи была пополнена данными база данных проекта ГИС "Палеолит Северной Азии" по археологическим материалам, геоморфологии и стратиграфии памятников Забайкалья. В результате на картах ГИС появился 51 археологический памятник (см. рис.1):

- | | | |
|--------------------|---------------------|-----------------------|
| 1. Хотык | 18. Фомичево | 35. Каменный ключ |
| 2. Каштак | 19. Кокорино | 36. Пойменная |
| 3. Санный мыс | 20. Ньянги 1 | 37. Куналейский мост. |
| 4. Ганзурино | 21. Боргой Сельгир | 38. Харенга |
| 5. Херексурин-Ури | 22. Белоозерск | 39. Кукун |
| 6. Крыжевский пост | 23. Халяша | 40. Ильмовая |
| 7. Дюрбены | 24. Горки | 41. Заречье 2 |
| 8. 314 км | 25. Береговая | 42. Базино |
| 9. Ивашка | 26. Окино ключи | 43. "589" |
| 10. Зарубино | 27. Шара-Горхон | 44. Верхний Мангиртуй |
| 11. Дэдэ-Ичетуй | 28. Кирпичная Падь | 45. Хотогой-Хабсагай |
| 12. Ньянги 2 | 29. Усть-Сахюрта | 46. Резаново |
| 13. Кандабаево | 30. Баин шки | 47. Ветрово 1 |
| 14. Черемушки | 31. Иркана | 48. Абайкино |
| 15. Читкан | 32. Удунга ий | 49. Димино |
| 16. Гремячий | 33. Малый сельгир | 50. Усть-Итанца 1 |
| 17. Мостовое | 34. Большой Сельгир | 51. Засухино |

В структуре баз данных ГИС (по тем археологическим объектам, по которым в опубликованных источниках представлены соответствующие материалы) выделены базовые показатели, признаки, характеристики и другие формы данных по археологическим памятникам (см. рис. 2), необходимые для их представления на HTML-страницах ГИС:

1. Общий вид.
2. История исследования.
3. Геоморфология

4. Стратиграфия.
5. Планиграфия.
6. Палинология.
7. Палеонтология.
8. Технология.
9. Типология.
10. Библиография.
11. Привязка к временным периодам:

- нижний палеолит
- средний палеолит
- верхний палеолит

12. Пещерные памятники

13. Памятники открытого типа:

- мастерские
- стоянки-мастерские
- стоянки
- местонахождение.

Информационную основу баз данных по перечисленным археологическим объектам составили результаты следующих информационно-статистических исследований:

- 1) анализ фракционного состава:

- нуклеусы и нуклевидные обломки
- орудия
- сколы
- гальки

- 2) группировочный анализ сколов

- 3) анализ и группирование объектов по видам орудий:

- леваллуа:
 - гладкие
 - двухгранные
 - фасетированные
- нелеваллуа:
 - пластины (гладкие, двухгранные, фасетированные, точечные)
 - отщепы (гладкие, двухгранные, фасетированные, точечные)

При разработке программно-технических решений по созданию ГИС авторами проекта учитывались следующие из требований, подготовленных и обоснованных в прешествующем году, которым должна геоинформационная система удовлетворять после завершения разработки:

- открытый доступ по WWW к накопленным материалам и результатам исследований в области палеолита Северной Азии;

- программно-информационное обеспечение ГИС на основе современных информационных технологий;

- обеспечение открытости и оперативности при внесении новых данных и обновления существующих информационных ресурсов ГИС.

При разработке цифровой карты памятников палеолита Северной Азии применялись технологии, представленные в открытых публикациях. При этом были использованы современные программные средства и утилиты, в том числе настольная геоинформационная система ArcView, а также утилиты собственной разработки. Для этой цели применялся метод растровой векторизации бумажного картографического материала.

При подобной векторизации данные (которые были сформированные в процессе работы над проектом реляционной базы данных палеолитических памятников, содержащей координаты «широта-долгота») была конвертированы в рабочий векторный формат Shapefile.

В качестве топографической основы были использованы свободно-доступные векторные карты мира и Российской Федерации, а также глобальная растровая географическая карта. Полученный в результате компоновки тематических и общегеографических слоев набор данных позволяет достаточно наглядно отобразить различные атрибутивные характеристики памятников палеолита Северной Азии, их географическое положение, периодизацию и общие показатели, характеризующие пространственно-временную распределенность археологического материала.

Для большинства географических объектов тематического слоя карты координаты заданы с точностью до 0.1 градуса. Точность полученной карты памятников достаточна для решения поставленных задач. Для доступа к ГИС-системе через Интернет используются средства сетевой ГИС MapServer. Для визуализации цифровых карт в режиме просмотра и редактирования используется ArcView GIS. При этом широко используются средства просмотра карт и редактирования векторных объектов. Была также апробирована технология ГИС TimeMap, позволяющая отображать на карте временную локализацию географических объектов, позволяющая отобразить на карте, например, памятники, имеющие датирование от 10 до 3 тыс. лет.

На основе базовых картографических слоев, свободно доступных в рамках международного проекта ECAI и тематических слоев, созданных в рамках настоящего проекта, разработана цифровая карта памятников палеолита, отражающая их периодизацию.

Сотрудничество с участниками международного проекта Electronic Cultural Atlas Initiative, направленного на решение задач по визуализации и анализа культурных и исторических феноменов средствами сетевой ГИС, позволило придать программно-техническим решениям настоящего проекта современный уровень представления. Созданный набор данных был зарегистрирован в проекте ECAI и является одним из немногих проектов, посвященных исследованию палеолитических культур.

В процессе работы над проектом применялись следующие аппаратные средства. Сервер Dual Pentium III под управлением операционной системы MS Windows 2000 Advanced Server является основным сервером проекта. На нем размещена реляционная база данных, программная часть сетевой ГИС, веб сервер сайта проекта, а также хранятся геоинформационные данные и атрибутивные данные. Для наполнения базы данных, создания векторных цифровых карт применялись рабочие станции на базе процессора Pentium III под управлением MS Windows 2000.

Подготовка цифровых картографических данных проекта производилась с помощью настольной ГИС ArcView. Данная ГИС позволяет создавать и редактировать векторные слои в формате Shapefile – рабочем формате пространственных данных проекта. ArcView также является основной ГИС-системой, применяемой на рабочих станциях локальными пользователями для работы с ГИС-ресурсами проекта. В ГИС ArcView был создан набор тематических векторных слоев, слой палеолитических памятников Сибири и Дальнего Востока, а также векторных слой памятников с атрибутивными данными, отражающими датировку памятников.

Доступ к картографическим данным через сеть Интернет организован средствами двух сетевых ГИС: Mapserver и Timemap. Обе они предоставляют пользователю возможность просмотра векторных карт проекта. С точки зрения технологии они представляют два различных метода генерации изображения карты. Для просмотра цифровой карты в Timemap применяется технология Java-апплетов. Это накладывает дополнительные требования к программному обеспечению клиента. Другой особенностью этой технологии является необходимость загрузки с сервера по сети векторных данных. Сетевая ГИС Mapserver является CGI-приложением, выполняемым на стороне сервера. Сгенерированное им изображение карты является обычным GIF-изображением, просмотр которого возможен средствами большинства веб браузеров. Возможности интерактивной страницы по просмотру карты был в процессе разработки ограничены масштабированием и сдвигом. По сравнению с Mapserver, в программном средстве технологии TimeMap содержатся более широкие возможности, сравнимые с настольной ГИС, включая возможность просмотра атрибутивной информации,

ассоциированной с объектом на карте (в нашем случае, имеется возможность просмотра информации о периодизации памятника, его стратиграфии и данных радиоуглеродного датирования). Была также использована возможность привязки слоя палеолитических памятников по URL к веб-страницам, динамически генерируемым по запросам к реляционной СУБД (см. рис. 3). Таким образом на основе технологии TimeMap были объединены в единую информационную структуру географические данные ГИС и относящаяся к ним всевозможная атрибутивная информация.

В качестве основной рабочей СУБД, обеспечивающей хранение данных и обработку пользовательских запросов была использована Oracle 8i, обладающая высокой надежностью и производительностью. Была спроектирована база данных, состоящая из восьми таблиц, хранящая информацию о названии, географических координатах памятника, типологию каменных орудий, орудийные наборы, найденные на памятнике в процессе раскопок, сведения о стратиграфии, данные радиоуглеродного датирования, общую информацию о памятнике в том числе историю исследования памятника и библиографию.

Доступ к базе данных на уровне администратора осуществляется при помощи программы SQL+, обрабатывающей запросы, построенные на языке SQL.

Для доступа к базе пользователей и редакторов, средствами технологии Enterprise Java Beans, языков XSQL и XML были созданы программные компоненты для Oracle Containers for Java, обеспечивающие доступ к базе данных Oracle на чтение и запись через Интернет по http-протоколу через гипертекстовые формы.

Доступ к базе данных на запись авторизован, просмотр – свободный.

Перспективным средством интеграции археологических данных проекта "Палеолит Северной Азии" в международную научную информационную среду представляется язык разметки археологических данных ArchaeoML, разрабатываемый в университете Чикаго в рамках проекта XSTAR.

XSTAR представляет собой средство интеграции данных различных информационных систем по археологии, также разрабатываются программные средства для анализа данных. Предложенные XSL-схемы языка ARchaeoML позволяют экспортировать атрибутивные данные проекта в общий открытый формат XML-документов.

Имеется возможность конвертировать средствами языка XSL данные в формат ArchaeoML при помощи использованных для публикации данных проекта в сети Интернет программных средств, а именно языка XSQL.

Доступ пользователей к информационным ресурсам проекта осуществляется через веб-сайт проекта, размещенный на сервере под управлением Oracle HTTP Server и контейнера сервлетов Oracle Containers for Java.

Интерфейсы разделены на редакторский и пользовательский. Интерфейс редактора позволяет кроме просмотра базы данных осуществлять редактирование орудийных наборов и общей информации о памятнике (см. рис. 4).

Интерфейс пользователя является основным средством доступа и позволяет просматривать цифровую карту палеолитических памятников, атрибутивные данные о памятнике, а также строить запросы к базе данных и делать выборки (см. рис. 5).

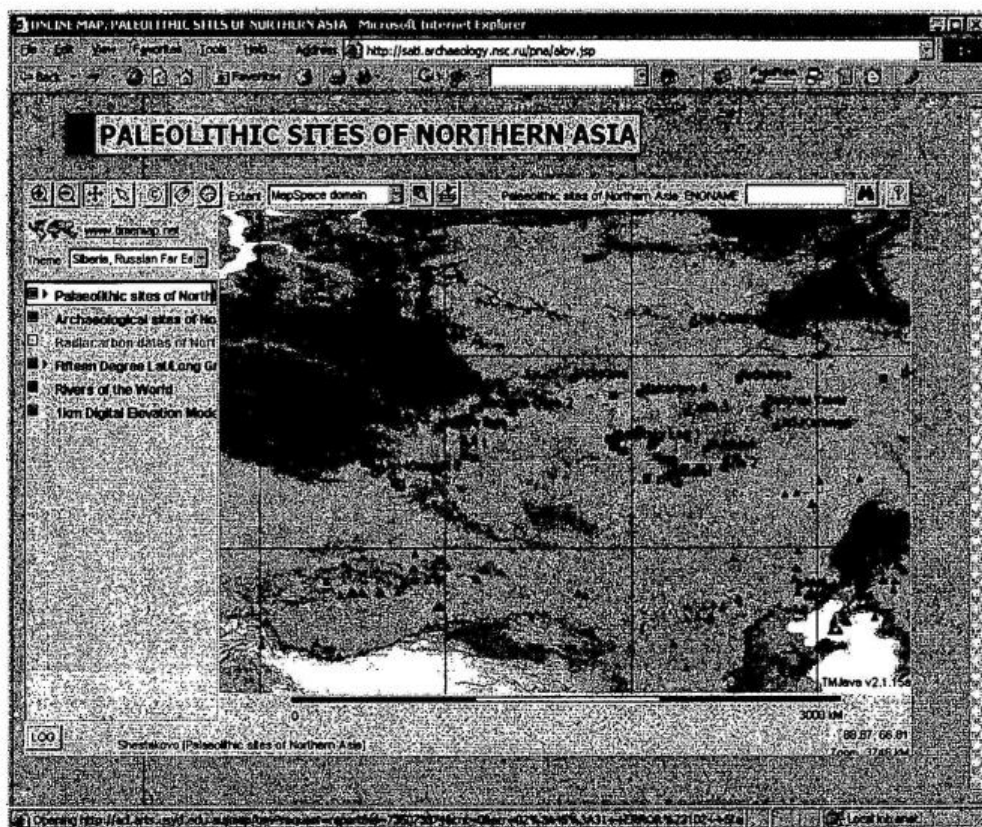


Рис. 1. Цифровая карта палеолитических памятников Северной Азии

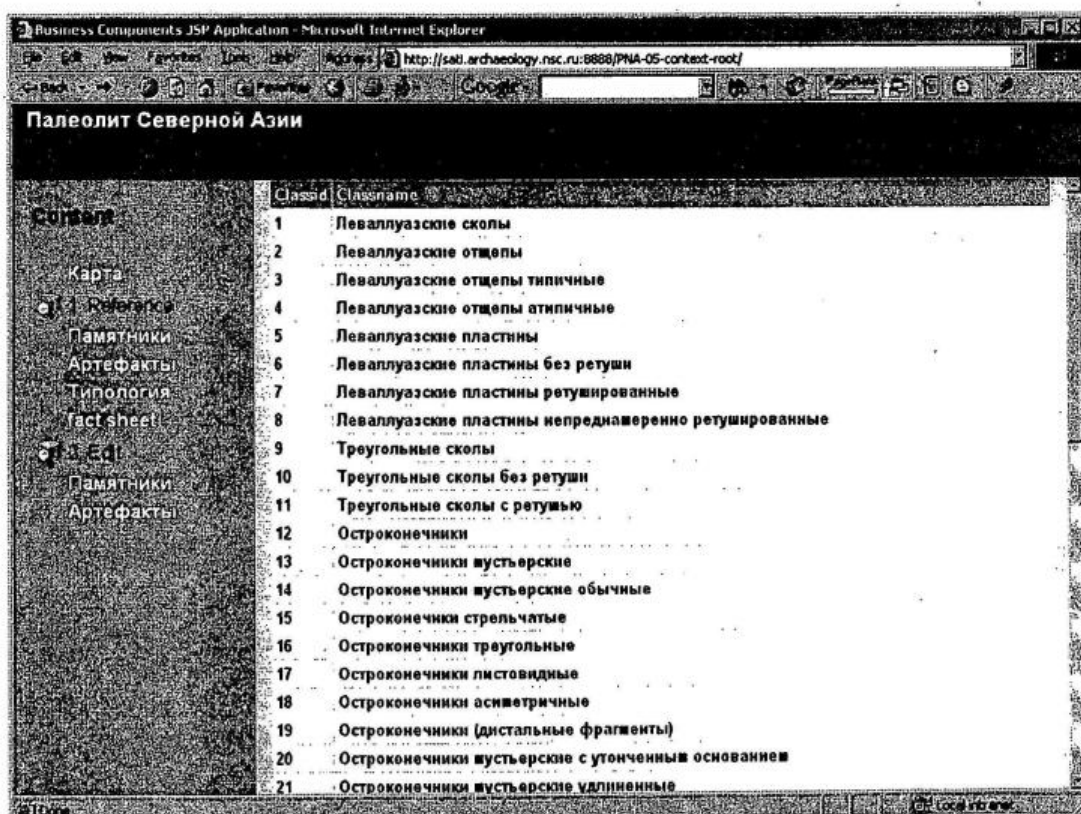


Рис. 2. Классификация каменных орудий.

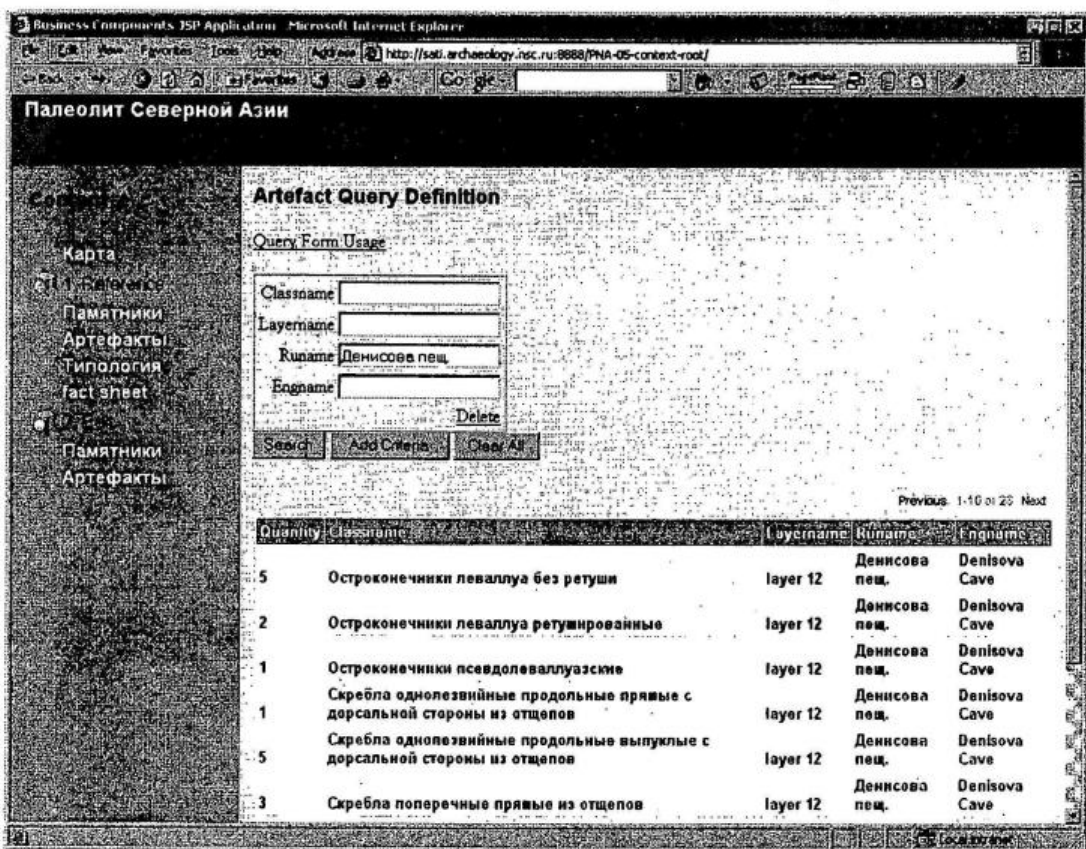


Рис. 3. Веб интерфейс к базе данных с примером поискового запроса.

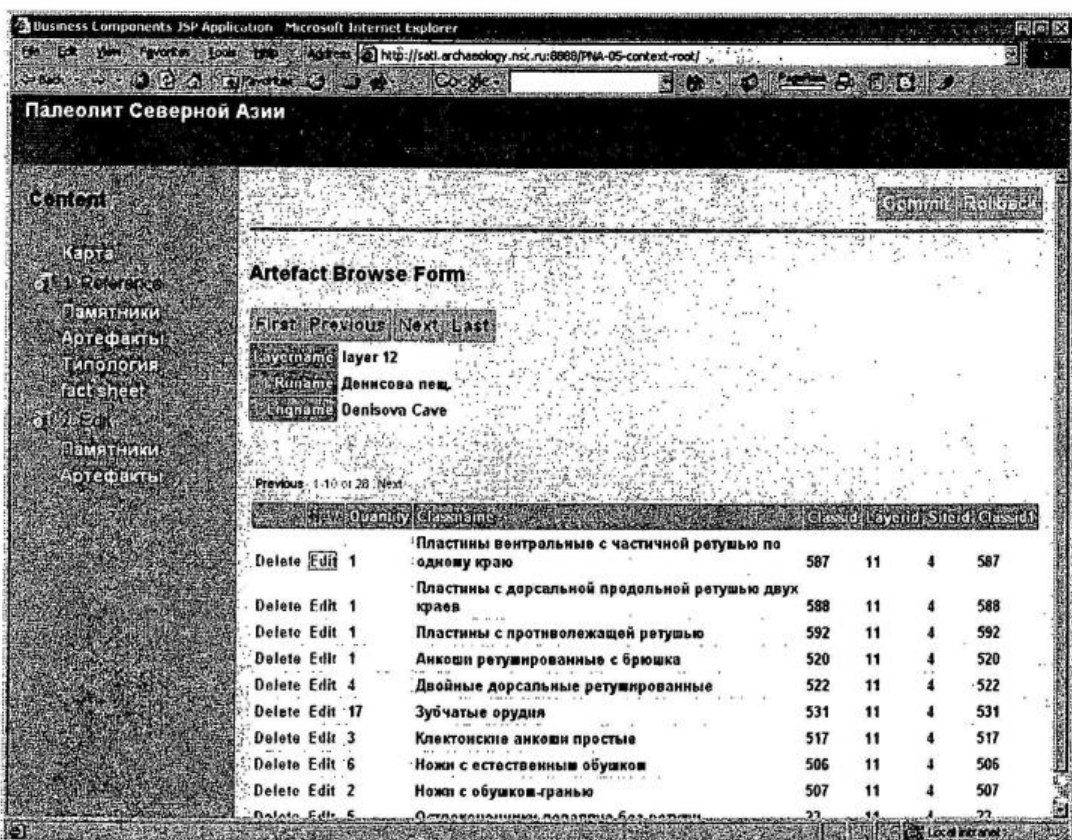


Рис. 4. Веб интерфейс редактирования базы данных.

PALEOLITHIC SITES OF NORTHERN ASIA		
Деревянка пещ. Demjova Cave		
query		
LATEPleistocene Paleolithic Sites		
site	pe	SOAN
	9000 +/- 95	SOAN-2743
	10690 +/- 65	SOAN-2862
	10800 +/- 40	SOAN-2863
layer 1	37235	SOAN-2864
layer 2	34700	SOAN-2488
layer 2	33390 +/- 1310	SOAN-2489
layer 2	35140 +/- 670	GX-17593
layer 2	46900 +/- 2300	GX-17602
main cave	9250 +/- 80	SOAN-3045
main cave	2495 +/- 190	SOAN-3044
main cave	8800 +/- 55	SOAN-3043

Рис. 5. Пример страницы с результатом запроса к базе данных.

Литература

- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Воробьев В.В., Федоров С.А., Бердников Е.В., Елагина С.В. Создание информационного центра сектора археологической теории и информатики ИАЭТ СО РАН (проблемы и решения) // Информационные технологии в гуманитарных исследованиях. Вып. 3. Новосибирск: Редакционно-издательский отдел НГУ, 2002. С. 6-14.
- Derev'anko A.P., Kholyushkin Y.P., Voronin V.T., Ekimov D.V., Goriatchev D.N., Schipunov V.V., Kopteva N.V. Concepts of informational and statistical processing of archaeological data in the computer center of the Institute of Archaeology and Ethnography in Novosibirsk. // Computer Applications and Quantitative Methods in Archaeology 1994 (BAR International Series 600), 1995: 203-206.
- Деревянко А. П., Холюшкин Ю.П., Ростовцев П.С., Воронин В.Т. Статистический анализ позднелептостеинских комплексов Северной Азии. Новосибирск: Изд-во НИИ МИОО НГУ, 1998
- Деревянко А. П., Холюшкин Ю. П., Ростовцев П. С., Воронин В. Т. Статистический анализ среднепалеолитических индустрий Ближнего и Среднего Востока. Новосибирск: Изд-во НИИУ ИДМИ, 1999.
- Холюшкин Ю. П., Бердников Е. В. Разработка цифровых карт палеолитических памятников Азии // Наследие древних и традиционных культур Северной и Центральной Азии. Материалы региональной археолого-этнографической студенческой конференции. Том 2. Новосибирск: НГУ, 2000. С. 35-38.
- Бердников Е. В. О проекте системы визуализации археологических памятников Северной Азии// Информационные технологии в гуманитарных исследованиях. Новосибирск: Изд-во НИИУ ИДМИ, 2000. С. 25-29.

Деревянко А.П.
Холюшкин Ю.П.
Воронин В.Т.
Ростовцев П.С.
Костин В.С.
Корнюхин Ю.Г.

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ИЗУЧЕНИИ КАМЕННЫХ ИНДУСТРИЙ*

За последние 30 лет произошли фундаментальные изменения в современной археологии. "Они подняли серьезные вопросы, касающиеся состояния теоретических разработок в археологии, привлекли внимание к стандартизации археологических наблюдений, совершенствованию методик сравнения, проверки археологических гипотез, палеоэкологических наблюдений. Все это потребовало внедрения методов различных наук в археологические исследования. Результаты использования одного лишь радиоизотопного анализа показали, что "посторонние" науки способны внести вклад в археологию в виде новой важной информации. Не менее важное значение имели совместные работы Брэйнерда и Робинсона, показавшие возможность математически решать археологические задачи" (Деревянко, Холюшкин, Фелингер, 1989: 10).

"В большинстве случаев целью исследования является выявление и установление закономерностей, тенденций, структурных соотношений между исходными объектами" (Деревянко, Холюшкин, Фелингер, 1989:134). Здесь и далее под объектом мы понимаем любой классифицируемый предмет, явление, процесс обладающий некоторыми признаками, следовательно обрабатываемый математически. Основным видом закономерностей между признаками является отношение зависимости их друг от друга (понимаемое в самом широком смысле). В операционном (математическом) отношении различают два основных вида зависимости: строго детерминированную (например, функциональную) и слабо или недетерминированную. Последняя как раз наиболее интересна для археолога-исследователя, так как большинство материала, с которым он имеет дело, по своей природе стохастично. Это происходит не только из-за ошибок измерения, отсутствия информации, неполноты знания законов природы, но и потому, что "многим природным и социальным явлениям внутренне присущи элементы стохастичности" (Деревянко, Холюшкин, Фелингер, 1989: 133). Одним из наиболее развитых аппаратов в математике на настоящее время для анализа слабодетерминированных зависимостей являются методы математической статистики. Основной идеей всех этих методов является определение в каждый конкретный момент на доступном материале степень зависимости (детерминированности) объектов друг от друга. Так 1 - означает полную (детерминированную) зависимость, которую подчас можно выразить функционально, а 0 - полную независимость объектов исследования друг от друга (или показатель того, что данного исходного материала недостаточно, чтобы применять к нему статистические методы исследования).

Идея создания собственных программных средств по информационно-статистической обработке данных археологических и науковедческих исследований для персональных ЭВМ появилась в 1989 году, когда в институте появились первые персональные IBM совместимые компьютеры. До этой поры А.Ф.Фелингером и В.Т.Ворониным был создан прообраз подобной программы, разработанной на алгоритмическом языке PL/1 для ЕС ЭВМ и предназначенной для эксплуатации в пакетном режиме. Основная процедура этой программы (группирование объектов) базировалась преимущественно на использовании энтропийно-информационного подхода А.Ф.Фелингера (Фелингер, 1985). В основу алгоритма положен критерий минимальности потерь информации при подборе кандидатов для образования групп.

Дополнительно к процедуре группирования в описываемую программу был включен блок информационно-статистического анализа исходной, промежуточной и результирующей информации. В частности, анализом был предусмотрен расчет информационно-статистических характеристик и показателей, разработанных и обоснованных А.Ф.Фелингером (Фелингер, 1985): мера неопределенности (величина

* Работа выполнена при финансовой поддержке РФФИ (проекты № 01-07-90393, 03-06-80418).

энтропии) одной совокупности выделенных признаков объекта (или множества объектов) относительно другой, мера структурированности (объем информации) одной совокупности выделенных признаков объекта или множества объектов относительно другой, те же показатели в контексте условных зависимостей, а также мера зависимости одних признаков от других и т.д. Одним из ключевых критериев информационного анализа, предложенных А.Ф.Фелингером, является отнесение оценки его результатов для каждой совокупности объектов к одному из следующих вариантов зависимости: 1) зависимость между группами признаков является сильной, близко к функциональной; 2) зависимость между группами признаков является средней; 3) зависимость между группами признаков является слабой; 4) между группами признаков зависимости практически нет.

Для ускорения анализа и перебора вариантов были использованы встроенные в язык PL/I средства по битовой обработке данных. На основе этих средств были разработаны алгоритмы построения битового портрета матрицы данных, отражающего структуру каждой группы. С помощью битового портрета отслеживается процесс группирования и изменения в составе групп.

Описанный программный комплекс использовался А.П.Деревянко, А.Ф.Фелингером и Ю.П.Холушкиным для обработки палеолитических материалов, результаты которой послужили основой для подготовки и выпуску монографии "Методы информатики в археологии каменного века" (Деревянко, Фелингер, Холушкин, 1989). В частности, результаты пошагового анализа процесса группирования были использованы Ю.П.Холушкиным для построения дерева группирования вручную.

Для реализации этого же алгоритма на персональных ЭВМ потребовалась существенная переделка не только в плане адаптации кода к новой операционной системе. Неодходимо было переписать этот код на другой алгоритмический язык, ибо с точки зрения алгоритмов битовой обработки данных язык PL/I на персональных компьютерах в среде MS DOS к тому времени был реализован в усеченном виде. В разработке программы для персональных ЭВМ, кроме В.Т.Воронина и А.Ф.Фелингера, принимал участие также программист Д.В.Малев-Ланецкий. В качестве нового языка программирования был выбран язык С, как наиболее отработанный на этот момент среди языков высокого уровня для IBM совместимых персональных компьютеров.

Дополнительно к основному алгоритму группирования и информационного анализа в новой программе был разработан комплекс для графического представления результатов. Были предусмотрены две формы графики для визуального анализа данных: 1) графики таблично заданных функций; 2) двоичные деревья, фиксирующие результаты процесса группирования.

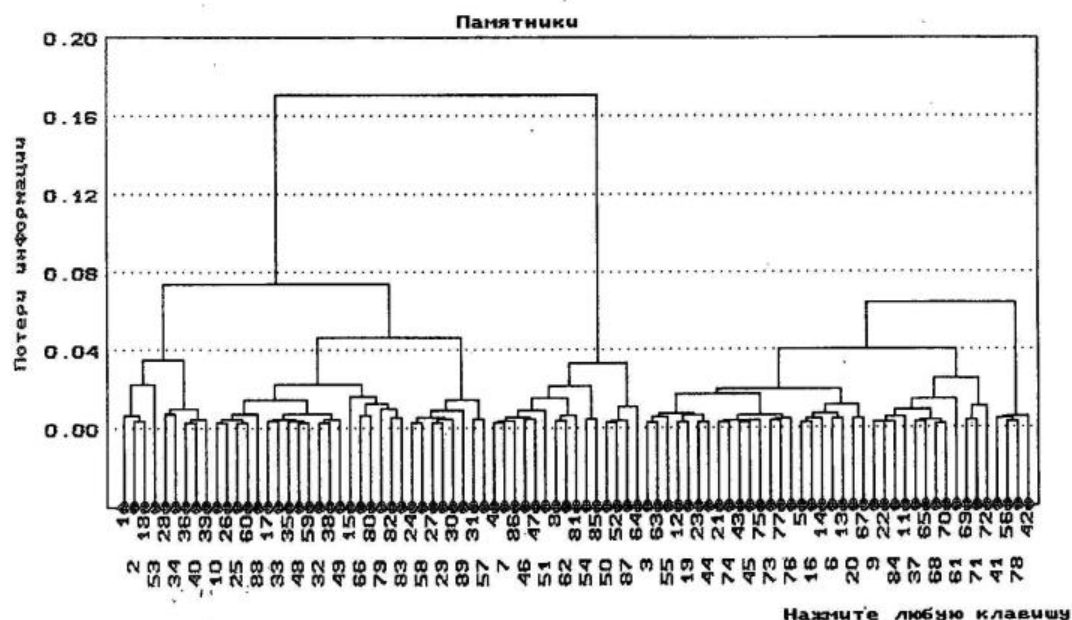


Рис. 1 Иллюстрация процесса группирования данных с помощью двоичного дерева.

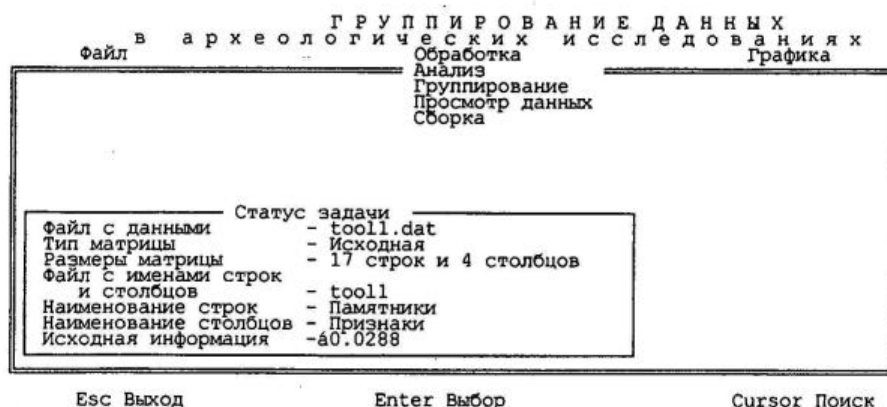


Рис. 2 Диалоговое окно программы группирования данных (вариант MS DOS).

Графики таблично заданных функций, будучи представленными в виде гладких кривых, имеют намного более наглядную форму в отличие от традиционных кусочно-линейных кривых, характерных для всех статистических пакетов того периода. Для интерполяции гладких кривых был выбран алгоритм локальных параболических сплайнов, сохраняющий выпуклость и монотонность (Воронин, 1987).

Для построения дерева, иллюстрирующего процесс и результаты группирования, в программе группирования фиксировались соответствующие "следы" этого процесса, отражающие последовательность объединения объектов в группы и потери информации, их сопровождающие. "Высота" ветвей определялась не только их размещением (относительным положением среди остальных ветвей на дереве), но и расстоянием от корня, т.е. величиной потерь информации при объединении (рис. 1).

Программный комплекс был выполнен в стандартной для программ, рассчитанных на выполнение под управлением MS DOS, манере, ориентированной на широкое использование диалога пользователя ЭВМ с помощью разнообразных меню (рис. 2).

Методы информатики, реализованные в описанной выше программе А.Ф.Фелингера, В.Т.Воронина, Д.В.Малева-Ланецкого, были использованы для обработки данных по мустьерским индустриям Алтая.

С момента образования в институте археологии и этнографии СО РАН неструктурной группы информатики к реализации методов кластерного анализа на персональных ЭВМ в среде MS DOS подключились В.В.Щипунов и Е.В.Коптева (Коптева, Щипунов, 1994). Ими была разработана программа по кластеризации объектов на основе евклидовой метрики и по детерминации - установлению функциональной зависимости одних признаков от других (Фелингер, 1985).

С 1996 года для обработки данных по мустьерским памятникам Северной Азии авторами - профессорами и преподавателями НГУ стали активно применяться методы и методики, встроенные в универсальный статистический пакет SPSS (версия 7.5), приобретенный Новосибирским государственным университетом при финансовой поддержке Института "Открытое Общество" (грант № 96-104-140696-1306).

Наряду с этим велась работа в содружестве с математиками программистами из Института Экономики и Организации Промышленного Производства СО РАН по обработке археологических комплексов с помощью разрабатываемого там программного комплекса.

В настоящей главе рассматриваются ряд методов типологического анализа археологических данных, разработанных под руководством П.С.Ростовцева.

Кластерный анализ как средство типологического анализа. Устойчивость кластеризации

В ходе историко-археологических исследований накоплено значительное количество материалов, которые необходимо систематизировать с целью выявления законов общественного развития, изучения культурной эволюции и совершенствования технологий. Эта работа требует от исследователя детального изучения данных и их обобщения, в ходе которого отдельные факты складываются в закономерности, а закономерности в теории.

Одним из методов математической обработки данных, полезным в этом отношении, является кластерный анализ. Кластерный анализ представляет собой средство исследования топологической структуры совокупности объектов. Он позволяет разбить множество объектов в признаковом пространстве на классы близких между собой объектов. Обнаруженные этим методом "сгустки объектов", называемые кластерами (классами, таксонами), позволяют сформулировать, в конечном счете, гипотезы о логической структуре совокупности.

В частности, этим методом можно изучать кластерную структуру множества памятников по наличию и частоте встречаемости артефактов, исследовать классификацию каменных орудий по их атрибутам, изучать структуру других совокупностей, представимых прямоугольными матрицами вида "объект-признак".

Мы будем считать синонимами термины "кластерный анализ", "автоматическая классификация", а также термины "классификация" и "разбиение".

Разнообразные виды кластерного анализа активно применялись и применяются в историко-археологических исследованиях (Жамбю, 1988; Деревянко, Фелингер, Холушкин, 1989; Cowgill, 1968; Деревянко, Холушкин, Воронин, 1994; Коптева, Щипунов, 1994 и др.).

Кластеры, замечательным образом найденные в первый раз и разумно описанные исследователем, после повторного сбора информации и применения кластерного анализа могут "рассыпаться" из-за случайности выявленной кластерной структуры.

Это разрушение может произойти тогда, когда реальная кластерная структура может отсутствовать вообще (исследуемая совокупность однородна) или когда задано не соответствующее реальности число классов. Первое по сути означает однокластерную структуру совокупности.

Достоверность наличия кластерной структуры обычно проверяется на основе привлечения дополнительных фактов, исследования классификации с использованием переменных, как участвующих, так и не участвующих в кластеризации. В данном случае практика в содержательной области исследования является основным критерием истины.

Точным математически обоснованным выводом о кластерной структуре уделяли внимание Бокк и Хартиган (Bock, 1985; Hartigan, 1985). Их исследования показывают сложность решения проблемы традиционными параметрическими методами.

Для решения задачи устойчивости мы остановимся на возможностях двух подходов, реализованных на основе статистического моделирования.

Первый подход представляет исследование, основанное на гипотезе отсутствия кластерной структуры данных. Изучается распределение остаточной дисперсии, полученной при кластеризации однокластерной совокупности - "идеально плоских для кластерного анализа данных". При работе с реальными данными изученное распределение и выборочная остаточная дисперсия могут служить основой для проверки гипотезы об отсутствии кластерной структуры. Здесь мы оценили, по существу, возможность применения классической теории проверки статистических гипотез.

Второй подход, на наш взгляд, более перспективный, основан на использовании метода Boot Strap (Efron, 1986) - имитации повторного сбора данных. Благодаря этому подходу, можно оценить как устойчивость классификации в целом, так и устойчивость классов и отдельных объектов. Ранее применение данного метода было разработано для исследования более сложных структур (Ростовцев и др., 1994).

Качество кластеризации.

Среди множества методов кластерного анализа мы рассматриваем лишь методы, позволяющие изучать большие массивы данных, не ограничивающиеся возможностями оперативной памяти ЭВМ. К таким методам относятся процедуры, разработанные Боллом и Холлом, МакКуином, алгоритм ИЗОДАТА (Дюран, Оделл, 1977). При этом будем считать, что критерий кластеризации - минимум разброса внутри кластеров.

При определении критерия кластеризации возможно использование различных метрик на множестве объектов, даже использующих неколичественные переменные, кроме того, в многомерном дисперсионном анализе используются более сложные выражения для разброса данных, связанные с применением детерминанта ковариационной матрицы (Петрович, Давидович, 1989). Практическое использование этого выражения для кластерного анализа затруднительно. Для определенности мы будем считать, что d - евклидово расстояние между векторами.

Пусть данные представляют собой матрицу типа "объект-признак" $[X_{ij}]$, $i = 1, \dots, m$, $j = 1, \dots, n$, где i - номер объекта, j - номер признака, элементы матрицы - значения признаков. Таким образом, объекты $\{X_i\}$ описываются как строки матрицы данных.

Пусть имеется разбиение совокупности объектов на k классов

$$R = \{R_1, \dots, R_k\}; R \in \{1, \dots, m\}; i = 1, \dots, k.$$

Общий квадратичный разброс совокупности объектов определяется формулой

$$SST = \sum_i d^2(x_i, \bar{x}), \quad (1.1)$$

где \bar{x} - центр тяжести всех объектов (среднее арифметическое соответствующих векторов).

Величина SST/m представляет собой обобщенную дисперсию многомерных данных.

Внутриклассовый разброс характеризуется показателем

$$SSQ(R) = \sum_i \sum_{R_i} d^2(x_i, \bar{x}_i), \quad (1.2)$$

где \bar{x}_i - центр тяжести R_i .

Величина $SSQ(R)/m$ представляет собой оценку остаточной дисперсии при объяснении разброса классификацией R .

Непосредственное использование $SSQ(R)$ неудобно для описания качества классификации, поскольку эта величина зависит от исходного уровня разброса данных SST . Гораздо удобнее применять долю остаточной дисперсии:

$$Q(R) = \frac{SSQ(R)}{SST}. \quad (1.3)$$

Может использоваться также доля объясненной дисперсии:

$$T(R) = 1 - Q(R). \quad (1.4)$$

Алгоритм кластеризации

Рассмотрим алгоритм, который последовательно строит разбиения совокупности объектов $R(2)$, $R(3)$, ..., $R(k)$ на два, три и более классов. При построении разбиений используется начальное разбиение, которое "улучшается" описанным ниже образом за счет перемещения объектов из класса в класс.

При построении $R(k)$ поочередно рассматриваются объекты X_i , $i = 1, \dots, m$; для каждого из них вычисляется величина $DSSQ(R, i, l)$ уменьшения $SSQ(R)$ при переносе этого объекта в класс l разбиения $R(k)$, при этом находится такой класс $l(i)$, для которого величина $DSSQ(R, i, l)$ максимальна. Объект переносится в найденный для него "лучший" класс $l(i)$, если при этом происходит улучшение критерия, и не переносится в противном случае. После просмотра всех объектов процесс повторяется. Работа алгоритма прекращается, если при очередной итерации не происходит улучшения критерия.

В качестве исходного разбиения $R(2) = \{R_1, R_2\}$ берется разбиение $\{\{\text{вся совокупность объектов}\}, \{\text{пустое множество}\}\}$.

В качестве исходного разбиения $R(k) = \{R_1, \dots, R_k\}$ рассматривается разбиение $R(k-1)$ с присоединенным к нему пустым классом R_k .

По существу этот алгоритм является модификацией известного алгоритма "к средних", который также является локально-оптимальным алгоритмом, минимизирующим $Q(R)$. Этот алгоритм состоит в следующем:

- 1) случайно выбираются k объектов - центров кластеров;
- 2) каждый объект относится к кластеру, расстояние до центра которого минимально;
- 3) вычисляются новые центры кластеров - средние арифметические по их объектам;
- 4) если хотя бы один объект переместился в другой кластер - осуществляется переход к п.1), в противном случае алгоритм заканчивает свою работу.

Заметим, что в оптимальной классификации алгоритм k средних не осуществит перемещения объектов.

Оценка значимости кластерной структуры на основе гипотезы идеально плохих данных.

Первый из изучаемых подходов основан на статистической гипотезе, состоящей в том, что данные представляют собой выборку из многомерной нормальной совокупности с

единичной ковариационной матрицей (данные идеально плохи). Такие данные представляют собой однокластерную структуру и кластеры в ней образуются только случайным образом. Для кластерного анализа, ввиду сложности алгоритмов, нет и в ближайшее время маловероятно появление необходимых формул, дающих функцию распределения оценки качества кластеризации (остаточной дисперсии). В данном случае, по-видимому, единственный путь - прямая оценка функции распределения функционала качества кластеризации путем проведения многочисленных экспериментов на случайных выборках из нашей "идеально плохой" для кластеризации совокупности.

Для сравнимости реальных данных с гипотетическими необходимо преобразование данных к шаровому облаку, соответствующему данным, имеющим единичную ковариационную матрицу.

Такая стандартизация данных проводится довольно просто: достаточно воспользоваться алгоритмом факторного анализа, дающим ортогональные факторы. Значения факторных оценок на объектах и будут стандартизованными данными. Можно также рассмотреть и другие методы стандартизации, связанные с нелинейной структурой расположения данных (Kruskal, 1964).

В дальнейшем мы предполагаем, что полученные результаты будут применяться к стандартизованным данным.

В ходе исследования данных целесообразно получать и исследовать классификации, последовательно увеличивая число классов, т.е. получая классификации с 2, 3, и т.д. классами.

Таким классификациям соответствуют значения долей необъясненной дисперсии $Q(2)$, $Q(3)$, ..., $Q(k)$.

Как предложили Д.Д.Холл и Д.Ханна (1986), в качестве сигнала о получении действительного числа кластеров может использоваться резкий переход от постоянного существенного уменьшения $Q(k)$ по сравнению с $Q(k-1)$ к незначительному их изменению.

Для выяснения, что же такое незначительные изменения $Q(k)$, в нашей работе исследуются относительные изменения функционала качества классификации:

$$q(2) = Q(2), q(3) = Q(3)/Q(2) \dots, q(k) = Q(k)/Q(k-1).$$

Очевидно, указанные функции качества разбиений Q и q связаны следующим соотношением из k классов:

$$Q(k) = \prod_{i=2}^k q(i). \quad (1.5)$$

Оценка распределения статистик изменения остаточной дисперсии.

Для оценки распределения $q(k)$ проведен ряд экспериментов. В каждом эксперименте:

- задавалась размерность матрицы данных - m и n ;
- исходя из многомерной нормальности генеральной совокупности, генерировалась матрица данных (Ермаков, Михайлов, 1982);
- проводилась кластеризация данных;
- результаты накапливались для статистической обработки (в виде объектов, соответствующих экспериментам, содержащим переменные m , n , $q(2)$, $q(3)$, $q(4)$ и т.д.).

Накопленный экспериментальный материал служит для оценки функции распределения долей необъясненной дисперсии $Q(k)$, а также функции распределения относительного уменьшения $Q(k)$ (значения $q(k)$) при переходе от $k-1$ -й классификации к k -й. Это позволяет, во-первых, проверить гипотезу H_0 - наличия в данных лишь однокластерной структуры против гипотезы H_1 - наличия в данных k - кластеров, во-вторых, проверить, похоже ли очередное уменьшение функционала на уменьшение, получаемое в однокластерных данных при попытке получить k классов.

Первое осуществляется следующим образом: на эмпирических данных получается классификация из k классов, вычисляется $Q(k)$ - доля остаточной дисперсии на эмпирических данных, а также вероятность

$$a(k) = \text{Prob}\{Q_{\text{выб}}(k) < Q(k)\} \quad (1.6)$$

случайно на однокластерной совокупности получить более качественную классификацию. Величина $p(k) = a(k)$ называется, также наблюдаемой значимостью. Если

эта величина мала (общеприняты уровни значимости 0.05, 0.01, 0.005, 0.001), гипотеза о наличии k кластеров не отвергается.

В принципе значение $a(k)$ может использоваться, также для определения действительного числа классов (Ростовцев, 1982). Перебирая $a(k)$, $k = 2, 3, \dots$, можно остановиться на минимальном значении $a(k)$ и соответствующее значение k принять за реальное число кластеров. Это означает, что мы выбираем классификацию, имеющую такое значение функции качества, меньше которого наименее вероятно получить на случайных однокластерных данных.

Величина $p = a(k)$ после такого перебора уже не будет соответствовать значимости наличия кластерной структуры, так как это - величина, найденная в результате перебора случайных величин; здесь мы сталкиваемся с проблемой множественных сравнений (Шеффе, 1963). Это не означает, что этими величинами нельзя пользоваться вообще - в подобной же ситуации успешно используются характеристики R -включения и R -исключения в пошаговом регрессионном анализе (Berk, 1977).

К сожалению, во многих случаях значение величины $a(k)$ бывает меньше погрешности, получаемой при ее вычислении. Сравнение величин $q(k)$ и его выборочного значения $q_{\text{выб}}(k)$ (соответствующих относительным приростам функции качества на случайных и эмпирических данных) в такой ситуации может оказаться более полезным. Здесь также рассматриваются вероятности:

$$b(k) = \text{prob}\{q(k) > q_{\text{выб}}(k)\} \quad (1.7)$$

Целесообразно остановиться на k классах, если величина $b(k)$ достаточно велика, а значит - случайно на однокластерных данных можно достаточно часто получить большее относительное уменьшение остаточной дисперсии. Иными словами, стоит прекратить работу, когда качество кластеризации достигает уровня, получаемого на "идеально плохих данных".

Возможно волнообразное поведение $b(k)$, поэтому не следует останавливаться на первом же локальном минимуме $b(k)$.

Организация экспериментов.

Число переменных в экспериментах задавалось равным 2, 3, 5, 7, 10; для этих размерностей признакового пространства моделировались выборки из - 15, 25, 50, 100, 200 объектов. В экспериментах с 15, 25, 50 объектами было проведено по 200 экспериментов; в экспериментах с 100 и 200 объектами - по 100 серий экспериментов.

На выходе экспериментов: $q(k)$ - доля остаточной дисперсии, полученная при переходе от классификации из $k-1$ кластеров к классификации из k кластеров. В проводимых экспериментах мы получали максимум 10 кластеров.

Полученные данные записывались в файл; причем после каждого эксперимента файл данных пополнялся строкой, содержащей размерность генерируемых данных и последовательность относительных изменений критерия качества: $m, n, q(2), q(3), \dots, q(10)$.

Исследование взаимной связи $q(i)$.

Поскольку $Q(k) = \prod q(i)$ - доля общей дисперсии, факт взаимной независимости $q(i)$ дал бы возможность специально не изучать распределение $Q(k)$, а использовать данные о распределении $q(i)$.

Взаимосвязь $q(i)$ мы изучили, рассматривая соответствующие корреляционные таблицы. Использование коэффициента корреляции предполагает двумерное нормальное распределение переменных, однако, как заметил Ван-дер-Варден (Ван дер Варден, 1964), при большом числе наблюдений результаты достаточно устойчивы к форме распределения. Поэтому в нашем исследовании мы допустили некоторую некорректность и использовали технику вычислений, применимую к нормальным данным.

Таблица № 1.

Матрица коэффициентов корреляции $q(i)$, $m = 50, n = 7, 200$ экспериментов.

	$q(3)$	$q(4)$	$q(5)$	$q(6)$	$q(7)$	$q(8)$	$q(9)$	$q(10)$
$q(2)$.197	.140	.167	.007	.062	.012	.054	.029
$q(3)$	1.00	.084	.251	.095	.050	.044	.028	.073

q(4)	.084	1.00	.075	.068	.084	.167	.199	.162
q(5)	.251	.075	1.00	.278	.080	.055	.100	.026
q(6)	.095	.068	.278	1.000	.146	.023	.113	.063
q(7)	.050	.084	.080	.146	1.00	.191	.042	.073
q(8)	.044	.167	.055	.023	.191	1.00	.151	.060
q(9)	.028	.199	.100	.113	.042	.151	1.00	.120

По проведенным экспериментам можно составить 25 матриц корреляции переменных $q(2) - q(10)$. Не утомляя читателя излишне подробным анализом этих матриц, продемонстрируем лишь данные для $m = 50, n = 7$ (табл.1).

Следовало бы ожидать существенных отрицательных корреляций. Это следует из того, что алгоритм последовательно извлекает информацию, "стремясь" на каждом шаге своей работы "объяснить" возможно большую часть дисперсии, и если на одном шаге удалось объяснить большую долю дисперсии, на следующих шагах это сделать должно быть труднее.

В действительности сплошные значимые связи (с уровнем значимости 0.01) в приведенной таблице 1 не наблюдаются, их в наддиагональной части таблицы всего 7.

Тем не менее число отрицательных коэффициентов в этой части матрицы равно 22 (из 36). Вероятность получить хотя бы две значимые связи с уровнем значимости 0.01 в условиях гипотезы взаимной независимости (и двумерной нормальности соответствующих распределений) равна 0.05.

Таким образом, сопоставляя полученные результаты с тем, что могло бы быть в условиях независимости, видно, что несмотря на достаточно непредсказуемый разброс коэффициентов корреляции и их незначимость в дальнейшем не следует отказываться от исследования закономерностей их взаимосвязи.

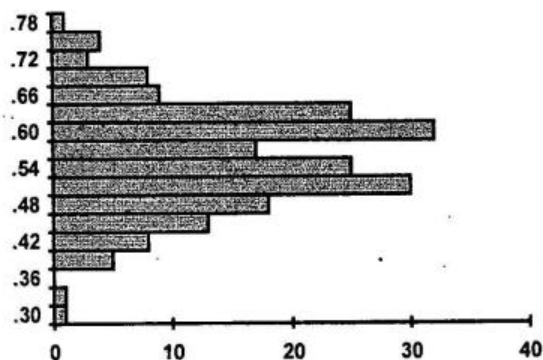


Рис. 3. Гистограмма распределения $q(2)$ ($m = 15, n = 2$).

Распределение $q(k)$.

Форма распределения $q(k)$ показана на рис. 3.

Изучение полученных гистограмм и сопоставление их с известными распределениями наводит на мысль, что мы имеем дело с бета-распределением, имеющем плотность распределения, равную

$$f_{ab}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \quad (1.8)$$

Оценки параметров a и b легко получаются на основе известных формул для среднего и дисперсии для этого распределения:

$$a = (M^2(x)(1 - M(x)) - D(x)M(x))/D(x), \quad (1.9)$$

$$b = a(1 - M(x))/M(x). \quad (1.10)$$

Оценки этих коэффициентов для данных различной размерности приведены в табл. 2.

Другой способ получения этих оценок - метод максимального правдоподобия. Функция правдоподобия для бэта-распределения, вычисляемая на основе выборки $\{x_1, \dots, x_N\}$ имеет вид:

$$L(a,b) = -\log\left(\prod f_{ab}(x_i)\right) = \quad (1.11)$$

$$= N \log(B(a,b)) - (a-1) \sum \log x_i - (b-1) \sum (1 - \log x_i).$$

Такой вид функции правдоподобия, а также аппроксимация производной гамма-функции, позволяет реализовать на основе метода "оврагов" алгоритм поиска максимума $L(a,b)$. Алгоритм достаточно трудоемкий, поэтому он был использован для контроля оценок a и b , полученных с использованием оценок среднего и дисперсии. Оказалось, что полученные двумя способами для части данных оценки практически не отличались. Развитие второго метода предполагается использовать для поиска зависимости параметров a и b от размерности исходной и генерируемой для кластерного анализа матрицы данных и числа классов.

Статистическая проверка гипотезы о бэта-распределении $q(k)$

Возможность найти параметры a и b , форма гистограммы, по виду похожая на график плотности бэта-распределения, еще не дают достаточных оснований утверждать о виде распределения исследуемых случайных величин. Для обоснования этого утверждения нами использовался критерий согласия "хи-квадрат".

Для его использования площадь под кривой плотности бэта-распределения разделяется интервалами на оси x на 10 равновеликих частей. Соответствующие границы интервалов $x_0 = 0, x_1 = 0.1, x_2, \dots, x_{10} = 1$ служат для интервализации значений q экспериментальных данных. При такой группировке наблюдений следует ожидать равных частот для каждого интервала. Таким образом, подтверждением теоретического бэта-распределения является незначимость отличия эмпирического распределения интервализованной переменной от равномерного.

Для получения границ интервалов дважды проводилось численное интегрирование функции

$$g(x) = x^{a-1}(1-x)^{b-1} \quad (1.12)$$

с шагом 0.0004, причем при вторичном интегрировании выяснялись границы, отсекающие десятые части полученного интеграла.

Таблица № 2. Оценки параметров распределения $q(k)$ при различных m, n, k

m	n	a b	Число классов									
			2	3	4	5	6	7	8	9	10	
15	2	a	18.3	7.6	6.2	3.8	3.3	3.2	3.0	2.8	2.2	
		b	14.5	3.8	2.2	1.1	0.9	0.8	0.8	0.8	0.8	
15	3	a	25.2	17.1	12.0	7.0	6.8	5.1	4.3	4.6	3.9	
		b	14.9	6.4	3.5	1.7	1.5	1.2	1.2	1.4	1.3	
15	5	a	59.4	34.5	20.7	18.6	14.5	10.9	9.1	8.1	7.7	
		b	21.9	9.3	4.9	3.7	3.0	2.5	2.4	2.0	2.5	
15	7	a	87.4	41.3	28.5	25.3	28.2	17.9	16.3	11.2	11.0	
		b	26.9	9.5	6.1	4.9	5.7	3.7	3.4	2.9	3.5	
15	10	a	134.6	65.7	45.7	39.3	32.5	29.2	22.4	23.3	19.1	
		b	33.0	13.0	8.0	7.1	5.9	5.5	4.4	5.5	5.3	
25	2	a	26.4	13.8	10.3	7.4	7.7	4.3	4.3	3.6	4.9	
		b	18.1	6.1	2.9	1.5	1.2	0.7	0.7	0.6	0.8	
25	3	a	52.3	28.4	25.4	14.4	12.4	10.6	9.1	6.5	7.3	
		b	24.0	9.0	5.5	2.6	1.9	1.6	1.2	1.0	1.1	
25	5	a	105.7	58.9	41.3	30.9	29.4	24.0	16.5	14.9	15.2	
		b	30.6	12.5	6.8	4.6	4.0	3.1	2.0	1.9	2.0	
25	7	a	174.9	85.1	59.0	45.3	43.0	35.8	28.2	27.4	22.6	
		b	40.8	14.7	8.4	6.2	4.9	4.3	3.2	3.0	2.6	

25	10	a	209.0	122.6	101.7	76.6	63.4	54.6	43.0	45.3	42.4
		b	39.6	16.7	11.9	8.6	6.4	5.7	4.4	4.8	4.4
50	2	a	67.7	24.5	19.1	12.7	13.2	11.1	9.8	5.9	6.4
		b	41.0	9.7	4.8	2.1	1.6	1.1	0.7	0.5	0.5
50	3	a	126.9	56.2	37.7	28.6	23.5	18.7	16.5	15.2	12.5
		b	48.1	14.8	7.1	4.0	2.7	1.7	1.5	1.3	1.0
50	5	a	199.6	100.6	82.8	71.1	62.2	39.1	31.4	26.9	38.7
		b	47.1	16.7	10.8	7.4	5.4	3.0	2.2	2.0	2.6
50	7	a	264.9	148.5	120.7	101.5	83.8	74.6	56.9	52.8	49.2
		b	48.9	19.0	12.2	8.6	6.2	5.0	3.8	3.3	2.9
50	10	a	429.3	312.2	213.3	168.1	136.1	121.6	95.2	86.9	89.2
		b	61.1	30.9	16.8	11.9	8.5	7.0	5.4	4.7	4.8
100	2	a	127.2	41.0	42.2	20.9	24.5	19.9	17.2	14.5	9.3
		b	69.4	16.8	10.3	3.8	2.9	1.7	1.2	0.8	0.4
100	3	a	204.8	97.8	62.1	49.2	45.4	39.6	25.9	28.6	16.9
		b	69.0	22.7	11.6	6.3	4.6	3.3	1.7	1.8	1.1
100	5	a	459.6	245.5	131.5	111.0	88.0	103.1	73.6	61.5	48.6
		b	93.9	34.5	13.9	9.4	5.9	6.4	3.9	3.0	2.5
100	7	a	503.3	414.7	249.2	238.6	168.1	173.2	96.7	79.2	100.0
		b	75.6	44.3	19.5	16.4	10.0	8.6	4.4	3.2	4.0
100	10	a	1080.3	608.0	416.5	290.2	319.2	216.0	213.3	242.8	166.7
		b	119.7	46.6	26.9	15.3	15.4	9.1	8.1	8.5	5.4
200	2	a	257.8	51.2	65.9	31.4	28.5	20.9	26.0	28.2	26.2
		b	135.1	20.9	14.6	4.6	3.3	2.0	1.7	1.3	1.0
200	3	a	490.8	220.7	121.6	91.1	71.8	63.4	56.5	47.6	44.4
		b	150.7	51.1	21.4	11.1	6.5	4.8	3.4	2.6	2.0
200	5	a	840.9	592.4	366.7	297.5	175.4	132.2	143.1	113.1	97.8
		b	151.4	76.7	36.1	23.1	11.2	7.2	7.1	5.0	3.7
200	7	a	1160.9	679.2	617.6	383.4	325.5	283.8	320.5	317.5	255.8
		b	152.0	62.1	45.3	22.6	16.4	12.2	12.4	10.8	7.8
200	10	a	1926.1	890.3	680.0	585.7	624.7	527.1	564.2	360.9	385.3
		b	184.6	59.8	37.0	25.0	24.0	18.0	16.2	9.6	9.6

В большинстве случаев проверка по критерию хи-квадрат показала незначимость расхождений распределений нашего теоретического и эмпирического распределений. Например, при $m = 15$, $n = 2$, $k = 2$ получено значение хи-квадрат, равное 10.7, что при 8 степенях свободы (число оцениваемых параметров) дает значимость $p = 0.219$; при $m = 50$, $n = 10$, $k = 10$ $p = 0.703$. Аналогичные результаты получаются практически для всех наугад взятых размерностей матрицы кластеризуемых данных.

Исследования части полученных данных показали, что величина $Q(k)$ тоже имеет распределение, близкое к бета-распределению. Этот факт также подтверждается на основе критерия хи-квадрат.

Таким образом, можно смело предположить, что для достаточно широкого диапазона размерностей матриц "идеально плохих" данных относительный прирост остаточной дисперсии при кластеризации данных имеет бета-распределение со своими параметрами a и b .

Статистическая проверка гипотезы о бета-распределении $q(k)$

Возможность найти параметры a и b , форма гистограммы, по виду похожая на график плотности бета-распределения, еще не дают достаточных оснований утверждать о виде распределения исследуемых случайных величин. Для обоснования этого утверждения нами использовался критерий согласия "хи-квадрат".

Для его использования площадь под кривой плотности бета-распределения разделяется интервалами на оси x на 10 равновеликих частей. Соответствующие границы интервалов $x_0 = 0$, $x_1 = 0.1$, $x_2, \dots, x_{10} = 1$ служат для интервализации значений q экспериментальных данных. При такой группировке наблюдений следует ожидать равных частот для каждого интервала. Таким образом, подтверждением теоретического бета-распределения является незначимость отличия эмпирического распределения интервализованной переменной от равномерного.

Для получения границ интервалов дважды проводилось численное интегрирование функции

$$g(x) = x^{a-1}(1-x)^{b-1} \quad (1.12)$$

с шагом 0.0004, причем при вторичном интегрировании выяснились границы, отсекающие десятые части полученного интеграла.

В большинстве случаев проверка по критерию хи-квадрат показала незначимость расхождений распределений нашего теоретического и эмпирического распределений. Например, при $m = 15$, $n = 2$, $k = 2$ получено значение хи-квадрат, равное 10.7, что при 8 степенях свободы (число оцениваемых параметров) дает значимость $p = 0.219$; при $m = 50$, $n = 10$, $k = 10$ $p = 0.703$. Аналогичные результаты получаются практически для всех наугад взятых размерностей матрицы кластеризуемых данных.

Исследования части полученных данных показали, что величина $Q(k)$ тоже имеет распределение, близкое к бета-распределению. Этот факт также подтверждается на основе критерия хи-квадрат.

Таким образом, можно смело предположить, что для достаточно широкого диапазона размерностей матриц "идеально плохих" данных относительный прирост остаточной дисперсии при кластеризации данных имеет бета-распределение со своими параметрами a и b .

Недостатки подхода.

Прежде, чем воспользоваться методом, необходимо провести большой объем вычислительной работы, создать индивидуальные таблицы для каждого алгоритма.

Необходима стандартизация эмпирических данных. Хотя стандартизация в большинстве случаев полезна, иногда она может увести нас от содержательной сути задачи.

Общая болезнь параметрических методов математической статистики - привязка к заданному гипотетическому распределению (в нашем случае многомерному нормальному).

Полученные в данном исследовании результаты позволяют рассматривать только количественные переменные.

Вычислительные сложности. Использование результатов исследования для относительно больших размерностей данных находится на грани вычислительных возможностей обычного программного обеспечения (в случае $m = 200$, $n = 10$ и $k = 2$ значения $a = 1926.1$, $b = 184.6$, а каковы же будут значения x^a ?!)

Исследование устойчивости имитацией повторного сбора данных.

Имитация повторного сбора данных предполагает, что наши данные представительны, т.е. их распределение близко к распределению генеральной совокупности. Для археологической информации это означает, например, предположение о том, что данные отражают структуру каменной индустрии в заданной пространственно-временной точке. Если исходить из этого предположения, то выборка с возвращением из совокупности объектов будет имитировать вторичный сбор данных.

Выборка с возвращением состоит в случайном выборе заданного числа объектов из данных и записи информации об этих признаках в файл, причем при каждом извлечении из данных объект остается в исходных данных, переписывается лишь информация о нем. Для полной имитации сбора информации объем этой выборки берется равным объему исходных данных.

Многократное вычисление различных статистик, оценок параметров на сгенерированных данных позволяет оценить вариабельность данных. Этот метод носит название метода Boot Strap. Благодаря его использованию удастся значительно упростить статистические выводы, весьма сложные задачи могут решаться просто и быстро.

Выборка с возвращением.

Выборка с возвращением приводит к новой, сгенерированной совокупности объектов, в которой часть объектов исходной совокупности может повториться несколько раз, а часть - отсутствует. При ее организации нет необходимости m раз (по числу объектов) обращаться к данным, извлекая очередной объект X_i . достаточно прочесть объекты последовательно, приписывая каждому объекту заранее сгенерированную кратность повторения объекта.

Исходная и "возмущенная кластерные структуры".

Допустим, мы провели кластерный анализ совокупности, получив исходную кластерную структуру. Это означает, что мы каждому объекту X_i приписали номер $g(i)$.

Поскольку объекты сгенерированной совокупности (выборки с возвращением) являются объектами исходной совокупности, можно считать, что сгенерированная совокупность расклассифицирована в соответствии с номерами $g(i)$. Эта классификация не будет оптимальной с точки зрения функции $SSQ(R)$, вычисленной по формуле (1.2).

Используя в качестве начального разбиения на сгенерированной совокупности эту классификацию, можно применить алгоритм кластерного анализа. Характерно, что кластерам исходной классификации $R = \{R_1, \dots, R_k\}$ соответствуют кластеры сгенерированной классификации $S = \{S_1, \dots, S_k\}$.

По тому, насколько изменится исходная кластерная структура, насколько не совпадают R и S , можно судить о ее устойчивости. Однако, как замечено выше, часть объектов исходных данных в сгенерированных данных отсутствует. Для проверки устойчивости этих объектов мы сталкиваемся с необходимостью перенести структуру кластеров сгенерированной совокупности на исходные данные - получить "возмущенную" кластерную структуру.

Полную информацию о кластерной структуре содержат центры кластеров. Поэтому обратный переход происходит следующим образом. В качестве центров кластеров возмущенной кластерной структуры возьмем центры классов S_1, \dots, S_k распределим объекты исходной совокупности в соответствии с методом k средних (шаг 2) по близости к центрам. Перенесенную таким образом на исходные данные кластерную структуру будем называть возмущенной кластерной структурой. В соответствии с этой классификацией каждый объект исходной совокупности X_i получит номер $s(i)$.

Статистики устойчивости.

Многочисленное повторение эксперимента по имитации сбора данных позволяет собрать информацию об устойчивости отдельных объектов, классов и классификации в целом. Пусть проведено F таких экспериментов.

Информация об устойчивости объекта.

Рассмотрим частоты $f_1(i), \dots, f_k(i)$, с которыми объект X_i оказывался в классах S_1, \dots, S_k соответственно ($f(i)$ - частота, с которой $s(i) = l$).

Величину $Pr(i, l) = f_l(i)/F$ назовем степенью предпочтения i -тым объектом l -го класса. Эта величина является, при сделанных предположениях о репрезентативности выборки данных, оценкой вероятности смены объектом X_i кластера при l , не совпадающим с $g(i)$, и вероятностью остаться в кластере при значении l , равном $g(i)$. Стабильность объекта, таким образом, характеризуется степенью предпочтения i -м объектом класса $g(i)$ (класса исходной классификации) - $Pr(i, g(i))$.

С точки зрения археологии, предпочтения $Pr(i, l)$ отражают устойчивость классификации отдельных объектов и близость их к другим кластерам.

Стабильность классов.

Обозначим F_l - общее число переходов объектов R_l в S_l во всех экспериментах.

Матрица $\|F_{lt}\|$, ($l = 1, \dots, k$; $t = 1, \dots, k$) по сути дела является таблицей сопряженности исходной (R) и "возмущенной" (S) классификаций вычисленной по множеству объектов - состояний перехода.

С ней можно обращаться точно также, как с обычной матрицей сопряженности признаков. В частности, можно проверить по критерию хи-квадрат гипотезу о независимости классификаций R и S . Таким образом, статистика хи-квадрат здесь является статистикой, по которой можно проверить гипотезу отсутствия кластерной структуры.

Доля переходов из класса R_l в класс S_t - $Pr_t(S_t/R_l) = F_{lt}/F_l$ - является оценкой вероятности соответствующего перехода, здесь F_l - сумма элементов матрицы $\|F_{lt}\|$ по строке (число исходов из R_l).

Эту величину можно также представить в виде среднего по F испытаниям долей объектов, совершивших такой переход. В историко-археологических исследованиях эта доля может показывать, какая в среднем часть выявленного кластерным анализом типа предметов древней культуры может быть присоединена к другому типу (при несовпадении номера исходного и "возмущенного" типа), насколько стабильно содержание типа (при совпадении l и t).

Стабильность/нестабильность классификации.

Стабильность $Stb(R)$ классификации в целом мы оцениваем долей переходов, оставивших объекты в исходном кластере. Величина $Stb(R)$ выражается формулой:

$$Stb(R) = \sum_i F_{ii} / (F_m) \quad (1.13)$$

Нестабильность, или оценка вероятности объектов не перейти в другой класс, оценивается по формуле:

$$Ust(R) = 1 - Stb(R) \quad (1.14)$$

В настоящее время еще не накоплен опыт в определении, какой уровень стабильности/нестабильности удовлетворителен, какой - неудовлетворителен. Традиции в данной области археологической науки сложатся в результате опыта обработки конкретных данных.

Доверительные интервалы - вторичные статистики.

Характеристики устойчивости объектов $Pr(i,l)$, классов $Pr(S_j/R_i)$ и классификации в целом $Stb(R)$ можно рассматривать с точки зрения наличия генеральной совокупности - результата бесконечного числа экспериментов, считая эти показатели оценками соответствующих параметров, полученных на выборке объема F .

При этом подходе целесообразно вычислить доверительные интервалы "истинных" значений параметров и стандартные ошибки их оценок.

Доверительные интервалы вероятности $p(i,l)$ объекта X_i случайно перейти в класс S_l получаются на основании статистики $f_l(i)$, имеющей биномиальное распределение с параметрами F и $p(i,l) - B(F, p(i,l))$. Оценкой $p(i,l)$ является $Pr(i,l)$, а ее стандартной ошибкой - $\sqrt{Pr(i,l)(1 - Pr(i,l)) / F}$.

Математическое ожидание $EStb(R_l, S_t)$ доли объектов класса R_l , перемещающихся в класс S_t оценивается средним значением - долей в каждом испытании - $Stb(R_l, S_t)$. Для оценки доверительных границ $EStb(R_l, S_t)$ следует воспользоваться приближенной нормальностью долей в каждом испытании (при большом объеме исходной совокупности) и распределением Стьюдента (попутно будет вычислена и стандартная ошибка оценки).

Аналогичным образом можно поступить, оценивая "истинное" значение нестабильности классификации в целом.

Недостаток подхода, другие возможности.

Основным недостатком метода является невозможность учесть объекты, которые по тем или иным причинам не попали в выборку. Этот недостаток в принципе исправляется применением "сглаженного метода Boot Strap", однако для его использования нужно иметь гипотезы о распределении отсутствующих данных. Формулирование этих гипотез в настоящее время пока не представляется возможным.

Подход можно обобщить на исследование устойчивости не только по сгенерированным данным, но и по отношению к исследованию структуры различных пространственно-временных и других срезов данных.

Алгоритмы анализа структуры и типологического группирования в анализе археологических данных

Информация, накопленная в ходе археологических исследований требует обобщения и анализа. Ее анализ без использования ЭВМ не может быть полным и охватит лишь

небольшую часть огромного фонда данных. Эффективно эту работу можно сделать во многих случаях, лишь используя специальные алгоритмы и программы анализа данных на ЭВМ.

Среди множества методов анализа данных в настоящем подпункте мы рассмотрим методы прямой кластеризации данных и автоматизации типологического анализа связи, которые, как нам представляется, могут оказаться полезными для исследования сложных и не полностью определенных историко-археологических данных.

Прямой кластерный анализ матриц данных развивался во многих работах отечественных и зарубежных авторов (Хартиган, 1975; Браверман и Мучник, 1983; Жамбю, 1988; Миркин, 1980; Ростовцев, 1982 ряд других авторов). Алгоритмы, описанные в этих работах затрагивали преимущественно структуры, основанные на прямоугольных блоках. В данной статье мы рассматриваем менее ограниченный класс структур, основанный на связных областях.

Типологическое исследование данных состоит в построении логических классификаций объектов. Эта задача также является достаточно распространенной в применении к естественно-научным и гуманитарным исследованиям (Лбов, 1981; Устинов и Фелингер 1973 и др.). Обычно рассматриваемые алгоритмы позволяют строить логические описания для детерминации классов разбиения совокупности объектов. В случае использования разнотипных переменных это описание ограничивается дихотомическим последовательным разбиением множества объектов. Здесь рассматривается логика последовательного разбиения (не обязательно дихотомического) с последующим объединением - "синтезом" типов. Целью построения классификации является минимизация остаточной обобщенной дисперсии по множеству переменных, в том числе и неколичественных. Работа является развитием разработок (Ростовцев, 1982), реализованных ранее на устаревшей в настоящее время технике.

При использовании многих методов анализа данных сложно делать выводы о статистической устойчивости результатов, поскольку до сих пор для части методов такие статистические оценки устойчивости не получены, для остальных методов эти оценки очень сложны или основаны на жестких предположениях о теоретическом распределении переменных. Для получения достоверных результатов в нашей работе используется метод Boot Strap (Эфрон, 1986).

Программы, реализующие указанные алгоритмы, написаны на алгоритмическом языке С++ на IBM PC. Вклад соавторов в данную работу: Н.Ю.Смирнова - разработка программы вычисления таблиц сопряженности, подключение к ней анализа структуры таблиц, В.С.Костин - разработка программы типологического группирования, подпрограмм обмена данными, Ю.Г.Корнюхин - системное сопровождение, разработка наиболее сложных элементов программной среды, Ростовцев П.С. - научное руководство, разработка методов, разработка программы поиска структуры таблиц. Ряд элементов потребовал общих усилий, в которых сложно выделить основного исполнителя.

Процедура выявления структуры таблицы

Для понимания и установления природы и характера различий большое значение имеет исследование структурных характеристик палеолитических памятников.

Выявление структуры таблицы происходит в два этапа. Первый этап состоит в упорядочении строк и столбцов таким образом, чтобы мало отличающиеся между собой строки (столбцы) оказались рядом. Критерием качества такого упорядочения является сумма расстояний между соседними строками (столбцами) - эта сумма минимизируется.

В ряде случаев имеет смысл отказаться от упорядочений строк и (или) столбцов, например, если строки являются временными рядами или элементы таблицы связаны с участками территории и требуется выявить однородные по какому-либо показателю регионы.

Второй этап - это разбиение элементов матрицы на связные области. Здесь критерий качества - дисперсионный: минимизируется остаточная дисперсия, получаемая при замене элементов матрицы на средние для элементов в соответствующих областях.

Упорядочение строк и столбцов

Естественно считать хорошо структурированной таблицу, в которой не очень часто происходят скачки по величине значений соседних элементов. Поэтому для лучшей

структурированности таблицы целесообразно переставить строки и столбцы матриц так, чтобы расстояние между соседними строками, а также расстояние между соседними столбцами в сумме было небольшим. Благодаря такой перестановке строки, соответствующие памятникам, упорядочиваются по близости распределений по артефактам. Таким образом обеспечивается сравнительно "плавный" переход от одного типа памятников к другому. Аналогичная цель преследуется при перестановке столбцов артефактов.

На основании этих соображений критерием качества упорядочения строк следует взять сумму расстояний между строками.

Если отождествить строки матрицы с вершинами взвешенного графа, где вес ребра, соединяющего две вершины, совпадает с расстоянием между ними, то задача минимизации становится весьма похожей на общеизвестную задачу "коммивояжера" - поиска обхода вершин графа (гамильтонова контура) минимальной длины. Ее отличие лишь в том, что искомый путь не замкнут. Искусственное присоединение к графу "нулевой" вершины, равноудаленной от всех остальных вершин, превращает указанную задачу в точности в задачу коммивояжера. В этом случае первая и последняя строки таблицы будут связаны фиктивной вершиной.

Существующие методы решения задачи коммивояжера делятся на два класса:

методы, приводящие к полной оптимизации, но в худшем случае требующие полного перебора вариантов;

локально-оптимальные методы, не всегда приводящие к оптимуму.

Здесь предпочтение отдано методам, принадлежащим ко второму классу. Они основаны на последовательном улучшении некоторого произвольного первоначально выбранного порядка обхода вершин. Алгоритм состоит в том, что пара вершин меняются местами. В расчетах последовательно рассматриваются пары вершин и, если при смене их мест происходит уменьшение пути, они действительно меняются местами.

Для этих целей использован метод "вставки", при котором из пути исключается некоторая вершина и затем вставляется между другими вершинами. Процесс перемещения вершин происходит до тех пор, пока длина контура уменьшается. Проблема перестановки столбцов решается аналогично.

Разбиение таблицы на связные области

После того, как строки и столбцы упорядочены и таблица приняла непрерывный вид, она разбивается на однородные части так, чтобы эти части объясняли по возможности большую часть дисперсии элементов. Предварительно уточним: что такое связные области и каковы их основные свойства; какой требуется критерий разбиения таблицы на связные области; в чем состоит локально-оптимальный алгоритм поиска разбиения таблицы на связные области.

Определение связных областей таблицы соответствует определению связного подпространства в задачах элементарного районирования (Воронин, Градова, 1987).

Назовем областью произвольное подмножество элементов таблицы. Два элемента будем называть соприкасающимися, если они находятся рядом, притом в одной и той же строке или в одном и том же столбце. Таким образом, из числа соприкасающихся исключаются элементы, соседние по диагонали. Область называется связной, если для любой выбранной пары ее элементов можно выстроить такую последовательность элементов, в которой каждая пара членов-"соседей" по последовательности соприкасается, причем первым и соответственно последним членами являются выбранные элементы.

Связность свидетельствует о том, что любая пара элементов области взаимозависима хотя бы косвенным образом, т.е. можно построить внутри области цепочку "соприкасающихся" пар артефакт-памятник, характеризуемых близкими частотами, содержащую эту пару элементов. Две связные области называются соприкасающимися, если существует пара соприкасающихся элементов, один из которых принадлежит одной области, другой принадлежит другой. Очевидно, объединение соприкасающихся связных областей дает связную область. Разбиением таблицы на связные области называется множество непересекающихся связных областей, которые покрывают все элементы таблицы.

Пример образа такой таблицы приведен в четвертой главе.

Критерием качества разбиения предполагается величина остаточного разброса при аппроксимации таблицы средними значениями элементов областей. При поиске разбиения эта величина должна, естественно, минимизироваться. Построение начинается с разбиения на самые мелкие области - по одному элементу в каждой. Такое разбиение дает минимальный, нулевой остаточный разброс, однако это разбиение неприемлемо из-за сложности. Поэтому производится последовательное объединение соприкасающихся пар областей: вначале находится пара областей, слияние которых даст наименьший прирост критерия, затем следующая пара и т.д. Подобные агломеративные алгоритмы используются в кластерном анализе, в частности, при анализе изображений, в анализе геологических данных.

Алгоритм объединения может работать до тех пор, пока все элементы таблицы не объединятся в одну область; возможна также ситуация, когда на некотором шаге объединения полученные области невозможно объединить из-за имеющихся в таблице неопределенных клеток. Однако такое окончание процесса бесполезно. На современном уровне решения задачи наиболее целесообразным является задание исследователем определенного числа кластеров, соответствующего его представлениям о сложности таблицы. В случае неудовлетворительных результатов имеется возможность найти новую структуру с другим числом кластеров.

Формальный критерий для числа кластеров базируется на следующих эвристических соображениях. Целесообразно ли увеличивать число кластеров, если при переходе к большему на единицу числу кластеров прирост объясненного структурой разброса составляет менее среднего разброса в пересчете на один элемент таблицы?

По-видимому, в большинстве случаев новый кластер, не объясняющий и доли дисперсии, приходящейся на элемент таблицы, не имеет смысла. Соответственно, если при слиянии кластеров потеря объясненного разброса превышает такую долю, то подобную потерю следует считать "существенной". Разумеется, приведенный критерий формален, приоритет имеют содержательные соображения.

В реализациях локально-оптимальных алгоритмов кластерного анализа целесообразно не ограничиваться агрегированием, а дополнить алгоритм процедурой перемещения объектов из класса в класс. Именно этим заканчивается работа по выявлению структуры таблицы.

Кроме того не менее важным представляется рассмотрение взаимосвязей между артефактами.

Исходным материалом для анализа структуры этих связей служит обычная для статистических методов матрица "объект-признак", где роль объектов играют памятники, а признаков - артефакты. Здесь нами рассматриваются булевы значения признаков: отсутствие (0) - наличие (1) артефакта (хотя возможно и более полное представление).

Целью анализа матрицы является выявление попарной взаимосвязи артефактов. В качестве коэффициента связи предлагается использовать Z-статистику, нормирующую гипергеометрическое распределение. Характерно, что этот коэффициент в условиях независимости переменных имеет нормальное распределение с параметрами (0,1). Исходя из 5% одностороннего уровня значимости можно считать значимой величину статистики, большую 1.7 по абсолютной величине, что позволяет сформулировать гипотезу о наличии связи.

Здесь техника анализа таблицы связей совпадает с техникой, использованной на предыдущем шаге исследования, но дополняется проверкой устойчивости с использованием метода статистического моделирования Boot Strap.

При анализе попарных связей между артефактами не рассматриваются диагональные элементы, поэтому они отмечены пустыми клетками (таким образом помечаются неопределенные значения). Не подвергнутая структуризации таблица позволяет увидеть лишь отдельные значимые связи и требуется специальная работа для обобщения информации.

Интересен в практическом и теоретическом плане вопрос об устойчивости выявленных структур. Что произойдет при повторном сборе данных, если бы были обнаружены другие памятники?

Моделирование такой ситуации основывается на гипотезе о представительности выборки: предполагается, что многомерное распределение данных отражает реальное распределение в генеральной совокупности. Это предположение позволяет моделировать повторный сбор данных за счет повторной выборки с возвращением из исходных данных, объем которой ("число памятников") совпадает с исходным числом объектов. Данные о некоторых памятниках в нашем случае повторяются, а некоторые памятники не будут выбраны. На смоделированных данных проводится проверка устойчивости выявленной структуры. В частности, для упорядочения артефактов проверяется, нет ли необходимости переместить какой-либо артефакт на другое место.

Литература

- Воронин Ю.А. Теория классифицирования и ее приложения. - Новосибирск, 1985.
- Гвоздовер М.Д., Григорьев Г.П. О фаціальности в верхнем палеолите (по материалам Каменной Балки II). // КСИА. - М., 1975, вып. 141, с. 12.
- Гинзбург Э.Х., Горенштейн Н.М., Ранов В.А. Статистико-математическая обработка шести мустьерских памятников Средней Азии. // Палеолит Средней и Восточной Азии. - Новосибирск, 1980: 7-31.
- Деревянко А.П., Фелингер А.Ф., Холушкин Ю.П. Методы информатики в археологии каменного века. - Новосибирск, 1989.
- Деревянко А.П., Фелингер А.Ф., Холушкин Ю.П. Опыт статистического группирования раннепалеолитических комплексов Евразии и Африки // Комплексные исследования палеолитических объектов бассейна р. Ануй. - Новосибирск, 1990:165-188.
- Деревянко А.П., Холушкин Ю.П., Воронин В.Т. Предварительные результаты информационно-статистического анализа мустьерских индустрий Алтая. // Методология и методика археологических реконструкций. - Новосибирск, 1994.
- Деревянко А.П., Холушкин Ю.П., Воронин В.Т., Ростовцев П.С. Некоторые статистические подходы к оценке фаціальности позднего палеолита Енисея. // Методология и методика археологических реконструкций. - Новосибирск, 1994.
- Миркин Б.Г. Анализ качественных данных и структур. - М., 1980.
- Муратов В.М., Аутлев П.У. Среднехаджожское раннепалеолитическое поселение. // Материалы по археологии СССР. - Л., 1971 (палеолит и неолит СССР, том 6): 41-48.
- Ростовцев П.С. Статистическое согласование мер связи в анализе социально-экономической информации. // Экономика и математические методы, 1991, т. 27, вып. 1: 150-156.
- Фелингер А.Ф. Статистические алгоритмы в социологических исследованиях. - Новосибирск, 1985.
- Binford L.R., and Binford S.R. A Preliminary Analysis of Functional Variability in the Mousterian of Levallois Facies // American anthropologist - 1966, ' 68:238-295.
- Bordes F. Principes d'une methode d'etude des techniques de debitage et de typology du Paleolithique ancien et moyen // L'anthropologie, v. 54:19-34.
- Kendall M.G. A Course in multivariate analysis. - London, 1957.

Деревянко А.П.	ПРЕДВАРИТЕЛЬНЫЕ ДАННЫЕ ПО СТРУКТУРНОМУ АНАЛИЗУ ТЕХНОЛОГИЧЕСКИХ ИНДЕКСОВ МУСТЬЕРСКИХ КОМПЛЕКСОВ СРЕДНЕЙ АЗИИ И КАЗАХСТАНА
Холушкин Ю.П.	
Воронин В.Т.	
Костин В.С.	

Статистическое изучение мустьерских комплексов Средней Азии мы считаем необходимым предварить лишь кратким обзором состояния проблемы, поскольку сведения о процессе изучения, осмысления и интерпретации материалов эпохи можно найти в работах А.П.Окладникова (1966), А.П.Окладникова и В.А.Ранова (1963), В.А.Ранова (Ранов, 1965, 1968, 1971, 1972, 1988, 1990), Р.Х.Сулейманова (1972),

* Работа выполнена при финансовой поддержке РФФИ (проект № 01-07-90393).

Н.А.Береговой (1960, 1982), В.А.Ранова и С.А. Несмеянова (1973), З.А.Абрамовой (1984), Л.Б.Вишняцкого (1996).

Мустьерская культура является самым исследованным периодом древнекаменного века Средней Азии. Общее количество пунктов с находками мустьерского времени приближается в настоящее время к сотне. Здесь известны пещерные стоянки (Тешик-Таш, Аман-Кутан, Обирахмат, Ходжакент, Огзи-Кичик) и местонахождения, материалы которых залегают в пролювиальных галечниках (Кара-Бура), в лессах (Худжи) или на поверхности (Семиганч). Большая часть местонахождений представлена пунктами с единичными находками. Наиболее информативные памятники находятся на территории Таджикистана и Узбекистана. Однако многие из этих материалов остаются неопубликованными. Большие трудности для получения сравнительных характеристик вызывает отсутствие единых приемов технико-типологического анализа палеолитических памятников.

До сих пор открытым остается вопрос о хронологических границах мустьерских индустрий Средней Азии. Для обоснования возраста большинства памятников отсутствует естественнонаучные данные (фаунистические послойные определения, палинологические характеристики, абсолютные датировки).

Судя по изолированным находкам леваллуазская техника и изделия мустьерского типа фиксируются в лессовых почвенных разрезах после пятой почвы (рисс-бюрм), т.е. менее 130 тыс. лет тому назад. Одновременно вплоть до 3-ей палеопочвы (50-40 тыс. лет тому назад) продолжается развитие каратауской галечной культуры (Ранов, 1988:21-22).

Таким образом существующая ныне фактологическая база по мустьерской эпохе Средней Азии оставляет желать лучшего.

Не требует доказательств положение о том, что формирование представлений о происхождении и генезисе палеолитических индустрий Средней Азии происходило под влиянием теории Х.Мовиуса (1944) и положений С.Н.Замятнина (1951). Так в работах А.П.Окладникова прямо говорится о расхождении путей развития палеолитической культуры на западе и востоке Средней Азии (1966). В известной степени можно отметить связь теории Мовиуса с концепцией В.А.Ранова. Прежде всего это видится в выделении двух линий развития, начиная с раннего палеолита до эпохи раннего металла. Первая линия (группа А) связывается с памятниками, расположенными к западу от Средней Азии, в основном с культурами Ближнего Востока, а вторая (группа Б) с памятниками, лежащими на востоке (культуры восточно-азиатского типа) (Ранов, 1965: 97; 1968). Отголоски этого влияния можно проследить и в совместной работе Г.П.Григорьева и В.А.Ранова (1973). В дальнейшем под влиянием работ Ф.Борда и его последователей В.А.Ранов к выводу, что "леваллуа-мустьерская группа не является такой однородной, как это казалось до сих пор" (Ранов, Несмеянов, 1973: 21). Это нашло свое отражение в выделении им на территории Средней Азии ряда технических вариантов, или фаций: леваллуазской (Джар-Кутан, Ходжакент, Оби-Рахмат), леваллуа-мустьерской (Кайрак-Кумы, Тоссор, Худжи, Капчигай, Ферганские стоянки); горного мустье (Тешик-Таш, Огзи-Кичик, Кутурбулак, Семиганч), и мустьеро-соанской (Кара-Бура, Ак-Джар, Кухи-Пиес) (Ранов, 1968; Ранов, Несмеянов, 1973: 23-25). Кроме того, учитывая новые данные по палеолиту Узбекистана, В.А.Ранов не исключил возможности выделения зубчатого мустье.

Как это следует из позднейших работ В.А.Ранова (Ranov, 1984: 313; Ранов, 1988:5-6), положение о существовании двух самобытных, не преходящих друг в друга традиций сохраняется. Однако для объяснения сосуществования указанных линий развития В.А.Рановым начинает отдаваться предпочтение гипотезе, согласно которой на нормальное развитие автохтонной каратауской галечной культуры накладывается пришедшая извне, вероятно с Ближнего Востока, леваллуа-мустьерская культура (Ranov, Davis, 1979). Таким образом авторами предлагается в качестве основной объяснительной гипотезы модель замещения. По этой модели в Средней Азии произошла относительно быстрая замена позднего варианта эректуса - носителя каратауской культуры пришедшим с запада неандертальцем (Ранов, 1988: 23). Данная модель находится в некотором противоречии с высказанной еще в 1965 году

В.А.Рановым мыслью о длительном существовании двух линий развития палеолитических индустрий Средней Азии: ближневосточной и восточно-азиатской.

Р.Х.Сулейманов предложил иной путь локального членения мустье Средней Азии. Им на основе различий в технических традициях, приемах расщепления и наборе орудий были выделены две группы мустьерских памятников: зубчатая (Кульбулак) и Оби-Рахматская культуры (Сулейманов, 1972:101), однако в процессе рассуждений он добавил к ним группу стоянок-мастерских (Учтут, Иджонт) и по сути дела согласился с выделением мустьеро-соанской группы (Сулейманов, 1972:117). Таким образом, если говорить по существу, Р.Х.Сулейманов объединил на основе новых, определенных им критериев ту же группу памятников, которая выделялась В.А.Рановым под наименованием леваллуа-мустьерской.

Беглый взгляд из Прикарпатья на мустье этого региона бросила Л.И.Кулаковская, который однако позволил ей лишь отметить, что "мустье Средней Азии представляется не столь мозаичным, как к этому... привыкли в Европе" (Кулаковская, 1990:213).

Еще более определенное об. однородности среднеазиатского мустье высказался Л.Б.Вишняцкий (1996:169). Однако и он не удержался от подчеркивания своеобразия Кульбулак (Вишняцкий, 1996: 169).

Что касается Казахстана, то там по технико-типологическим показателям выделены: типичное мустье (Кошкурган), леваллуа-мустье (Актогай, Семизбугу (комплекс А и Б), мустье-соан (Музбель), зубчатое мустье (Бурма) и мустье с ашельской традицией фации леваллуа (Хантау, среднемустьерский комплекс Семизбугу) и нелеваллуа (позднемустьерский комплекс Семизбугу) (Деревянко, Маркин, Васильев, 1994: 201).

Подводя итоги развитию взглядов исследователей на характер палеолита Средней Азии можно отметить, что здесь, исследователи сталкиваются с проблемой как ламперов, так и сплиттеров.

Данная глава не охватывает всех проблем, связанных с изучением мустьерских комплексов Средней Азии. Перед нами стоит гораздо скромная задача - дать представление о возможностях, которыми располагает исследователь для структурного и сравнительного анализа технических, типологических индексов, основных категорий каменного инвентаря. Иными словами, нами предпринята попытка проведения технико-типологического количественного анализа на основе современных технических возможностей. Задачу анализа облегчает то обстоятельство, что почти все памятники в общем плане более или менее одновременны (развитое или позднее мустье). Значительная часть материалов была обработана одним из авторов монографии, а опубликованные материалы использовались из работ двух авторов - В.А.Ранова и О.А.Артюховой (табл. 1.).

Исходными данными для исследования структуры служит таблица 18 мустьерских комплексов Средней Азии (строки и пяти технологических индексов (столбцы), взятых из ряда публикаций по этому региону (Ранов, 1972; Артюхова, 1998а; 1998б).

Первоначально для улучшения структурированности таблицы были упорядочены строки и столбцы так, чтобы мало отличающиеся строки (столбцы) оказались рядом. Благодаря такой перестановке строки, соответствующие археологическим комплексам упорядочились по близости распределений технологических индексов. Аналогичная цель преследовалась при перестановке столбцов-индексов. Итоги подобных перестановок представлены в таблице 2.

В результате упорядочивающих перестановок строк и столбцов сумма расстояний между соседними столбцами уменьшилось на 18.4 %, а сумма расстояний между соседними строками уменьшилось на 35.3 % соответственно.

После упорядочения строк и столбцов было предпринято выявление связанных областей, позволяющих выделить группы значений технологических индексов, объединяемых по сходству связей со значениями археологических комплексов Средней Азии.

Таблица 1. Технологические индексы мустьерских комплексов Средней Азии (по материалам публикаций).

№ п/п	Памятники	Технологические индексы				
		IL	ILty	IF	Ifst	ILam
1	Худжи	14.5	87.0	27.0	6.0	62.0
2	Тоссор	12.2	78.0	25.6	4.4	20.6
3	Джар-кутан	37.8	85.5	55.0	41.0	53.0
4	Георгиевский Бугор	9.5	27.4	44.3	8.8	10.8
5	Кара-Бура	4.2	17.6	23.6	11.5	6.3
6	Семиганч	20.2	27.4	37.1	14.2	30.7
7	Огзи-Кичик	32.7	31.0	35.9	4.6	44.0
8	Ходжакент	44.0		33.0	5.0	26.0
9	Кайрак-Кумы				38.0	56.0
10	Чингиз	41.1	23.3	54.0	29.0	23.0
11	Кош-Курган	3.9	12.5	42.3	13.7	6.1
12	Актогай	44.0	61.5	48.6	8.6	6.6
13	Семизбугу 10 А	9.8	20.3	50.8	6.0	2.8
14	Семизбугу 10 В	9.8	29.4	57.6	15.2	2.8
15	Семизбугу 10 С	11.4	30.0	21.4	11.9	7.0
16	Семизбугу 10 D	6.8	13.2	14.3	4.8	4.1
17	Хантау	3.8	29.5	35.7	3.6	2.1
18	Бурма	0.8	3.7	26.8	9.8	2.1

Таблица 2. Упорядоченные данные

№ п/п	Памятники	Технологические индексы				
		IF	IL	IFst	ILam	ILty
1	Худжи	27.0	14.5	6.0	62.0	87.0
2	Тоссор	25.6	12.2	4.4	20.6	78.0
3	Семиганч	37.1	20.2	14.2	30.7	27.4
4	Огзи-Кичик	35.9	32.7	4.6	44.0	31.0
5	Ходжакент	33.0	44.0	5.0	26.0	..
6	Актогай	48.6	44.0	8.6	6.6	61.5
7	Джар-кутан	55.0	37.8	41.0	53.0	85.5
8	Кайрак-Кумы	38.0	56.0	..
9	Чингиз	54.0	41.1	29.0	23.0	23.3
10	Семизбугу_10_В	57.6	9.8	15.2	2.8	29.4
11	Семизбугу_10_А	50.8	9.8	6.0	2.8	20.3
12	Георгиевский Бугор	44.3	9.5	8.8	10.8	27.4
13	Хантау	35.7	3.8	3.6	2.1	29.5
14	Семизбугу_10_С	21.4	11.4	11.9	7.0	30.0
15	Кара-Бура	23.6	4.2	11.5	6.3	17.6
16	Семизбугу_10_D	14.3	6.8	4.8	4.1	13.2
17	Бурма	26.8	0.8	9.8	2.1	3.7
18	Кош-Курган	42.3	3.9	13.7	6.1	12.5

Возникает вопрос: насколько устойчиво и в какой мере репрезентативно это разбиение с точки зрения методологии статистического анализа.

Для выяснения этого вопроса при последующем анализе таблицы использовался алгоритм для выделения связанных областей, примененный нами ранее в сравнительных исследованиях мустьерских комплексов Ближнего и Среднего Востока и Кавказа.

Всего в матрице 86 элементов, разбиваем на 10 областей. Образованная этим разбиением кластерная структура наглядно представлена в таблице 4 и на рис. 1.

Как и раньше, линией обведены области, номера областей выделены жирными цифрами внутри областей. Здесь оттенком серого цвета выделены пустые ячейки. Для ориентации в расположении областей удобно воспользоваться схемой на рис. 1.

Таблица 3. Результаты "серого" анализа данных по технологическим индексам мустьерских комплексов Средней Азии..

Памятник	IF	IL	IFst	ILam	ILty
Худжи	27	14.5	6	62	87
Тоссор	25.4	12.2	4.4	20.6	78
Семиганч	37.1	20.2	14.2	30.7	27.4
Огзи-Кичик	35.9	12.7	4.6	44.0	31
Ходжакент	34	44	5	26.0	
Актогай	48.6	44	8.6	6.6	61.5
Джар-кутан	55	37.8	41	53.0	85.5
Кайрак-Кумы			38	56.0	
Чингиз	54	41.1	29	23.0	23.3
Семизбугу 10 В	57.6	9.8	15.2	2.8	29.4
Семизбугу 10 А	50.8	9.8	6	2.8	20.3
Георгиевский Бугор	44.5	9.5	8.8	10.8	27.4
Хантау	35.7	3.8	3.6	2.1	29.5
Семизбугу 10 С	21.4	11.4	11.9	7.0	30
Кара-Бура	31.6	4.2	11.5	6.3	17.6
Семизбугу 10 D	14.3	6.8	4.8	4.1	13.2
Бурма	26.8	0.8	9.8	2.1	3.7
Кош-Курган	42.5	3.9	13.7	6.1	12.5

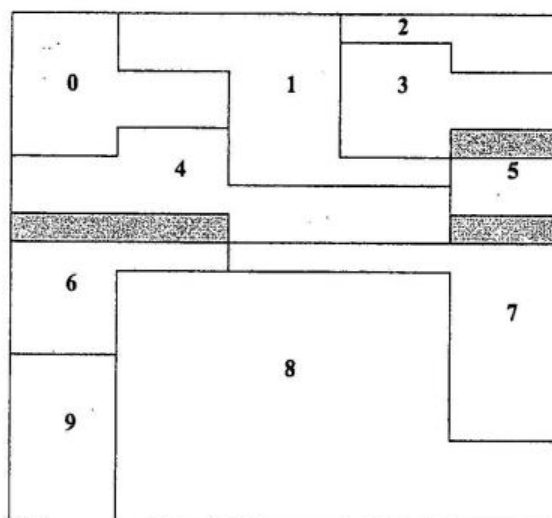


Рис. 1. Кластерная структура после объединения кластеров

На предварительной стадии исследования структуры упорядоченных данных была сделана попытка использовать "серый анализ", пометив эти данные оттенками серого цвета, исходя из величины значений в клетках таблицы:

- клетки белого цвета – нет данных,
- клетки светлосерого цвета – относительно малые значения в клетках,
- клетки темносерого цвета – значения индексов в клетках в пределах среднего диапазона,
- клетки черного цвета – большие значения индексов.

Результаты серого анализа представлены в таблице 3, оформленной в виде графического рисунка (см. табл. 3.)

На рисунке отчетливо просматриваются три области белого цвета, две области светло-серого цвета, три области темно-серого цвета и четыре области черного цвета. Каждая из этих областей составлена из попарно соседних клеток одного оттенка.

Здесь оттенком серого цвета выделены пустые ячейки. Нетрудно заметить, кластерная структура, зафиксированная на рис 1, во многом совпадает со структурой, выявленной с помощью "серого анализа" (см.табл. 3).

Таблица 4. Распределение элементов по 10 областям.

№ п/п	Памятники	Технологические индексы				
		IF	IL	IFst	ILam	ILty
1	Худжи	27.0	14.5	6.0	62.0	87.0
2	Тоссор	25.6	12.2	4.4	20.6	78.0
3	Семиганч	37.1	20.2	14.2	30.7	27.4
4	Огзи-Кичик	35.9	32.7	4.6	44.0	31.0
5	Ходжакент	33.0	44.0	5.0	26.0	
6	Актогай	48.6	44.0	8.6	6.6	61.5
7	Джар-кутан	55.0	37.8	41.0	53.0	85.5
8	Кайрак-Кумы			38.0	56.0	
9	Чингиз	54.0	41.1	29.0	23.0	23.3
10	Семизбугу_10_B	57.6	9.8	15.2	2.8	29.4
11	Семизбугу_10_A	50.8	9.8	6.0	2.8	20.3
12	Георгиевский_Бугор	44.3	9.5	8.8	10.8	27.4
13	Хантау	35.7	3.8	3.6	2.1	29.5
14	Семизбугу_10_C	21.4	11.4	11.9	7.0	30.0
15	Кара-Бура	23.6	4.2	11.5	6.3	17.6
16	Семизбугу_10_D	14.3	6.8	4.8	4.1	13.2
17	Бурма	26.8	0.8	9.8	2.1	3.7
18	Кош-Курган	42.3	3.9	13.7	6.1	12.5

Таблица 5. Дисперсионный анализ 10 областей.

Область	Среднее	Ср.кв.откл.	Объем	Объясняет долю дисперсии (в %)	
				область	элемент области
0	30.214	6.132	7	0.6	0.089
1	8.456	4.128	9	6.8	0.753
2	75.667	12.662	3	22.5	7.485
3	29.950	7.856	6	0.5	0.081
4	46.378	7.069	9	12.2	1.359
5	73.500	16.971	2	13.7	6.863
6	49.560	6.803	5	8.9	1.785
7	25.500	4.581	9	0.0	0.002
8	7.293	4.055	30	25.9	0.865
9	27.350	10.132	6	0.1	0.021

В таблице 5 представлены вклады областей в объяснение разброса элементов таблицы. В частности, наиболее существенными являются области 2, 4, 5, 8, существенными 1, 6. Среднеквадратичное отклонение по таблице = 20.23. Объясненная дисперсия составляет 91.3 %.

ТИПОЛОГИЧЕСКИЙ АНАЛИЗ ТЕХНОЛОГИЧЕСКИХ ИНДЕКСОВ

Типологией обычно называется логическое разбиение совокупности объектов на качественно различные группы объектов – типы. Построение типологий можно осуществлять чисто умозрительно, не имея наличия данных. Например, все артефакты можно разбить на односторонние и бифасиальные. Построение же типологий на основе эмпирических данных основывается на переменных, которыми эти данные описаны.

В соответствии с таким пониманием типологии в данном разделе нашей работы рассматриваются множества из 2 целевых переменных, по которым оценивается качество группирования. Основная идея группирования может быть выражена в рабочей формуле:

Типология = Логика группирования + Цель группирования

Целью группирования здесь является разделение совокупности объектов на классы, различающиеся по множеству "зависимых" переменных. Логика группирования определяет способ и критерий этого разделения. Оптимизация логики группирования состоит в максимизации меры связи между разбиением и совокупностью целевых переменных по множеству допустимых разбиений.

В типологии используются идеи, заложенные в методах кластерного анализа, основанного на дисперсионном анализе [Hartigan, 1975; Дюран, Оделл, 1977; Жамбю, 1988; Ростовцев, Костин, 1995]. Логика автоматического группирования имеет много общего с методами построения логических решающих правил [Лбов, 1981, 1983].

Построение таких типологий происходило в два этапа.

Первый этап – анализ – состоит в последовательном разбиении совокупности объектов (группировании) объектов по признакам. Прежде всего, по каждому из независимых признаков производится оптимальная с точки зрения критерия группировка объектов, и "лучшая" среди этих группировок берется в качестве начального приближения типологии. При этом для каждого признака рассчитывается доля объясненной группировкой дисперсии целевых (зависимых) переменных, что и служит основанием выбора "лучшего" разбиения.

На следующем шаге выбирается "оптимальная" с точки зрения критерия пара – один из классов полученного разбиения и один из признаков, по значениям которого группируются объекты этого класса, и получается некоторая группировка. На последующих шагах процедура повторяется.

Процесс идет до тех пор, пока исследователь не решит, что полученный результат удовлетворяет его по полноте описания связи систем переменных, либо дерево группирования достигает заранее заданного числа его вершин. Заранее, перед началом группирования, задается параметр ветвления – число групп, получаемых на шаге группирования по отдельной переменной. В нашем примере параметр ветвления задан равным двум – на каждом шаге группирования разбиваемая группа делится на две части.

Типологическое дерево представлено на рис. 2.

Здесь тип 1, группу 1 представляют памятники из Казахстана: Кош-Курган, Семизбугу 10 А, Семизбугу 10 В, Семизбугу 10 D, Хантау, Бурма, и один памятник из Таджикистана – Кара-Бура. Тип выделился по признаку «индекс пластинчатости» со значениями от 1.42 до 6.2.

Впоследствии к типу 1 была отнесен группа 7, в которую вошел комплекс Семизбугу 10 С.

Для типа один характерны низкие показатели индексов леваллуа и индекса леваллуа типологического (табл.6).

Тип 2 представлен группой четыре. Группа выделилась по высоким показателям индекса фасетирования, имеющего характерный для группы разброс значений от 48.6 до 55.0. Отличительной чертой группы являются очень высокие индексы леваллуа и леваллуа типологического (табл.6). В группу вошли: Чингиз, Актогай и Джар-Кутан.

Тип 3 представлен группой 2 из трех комплексов: Георгиевский бугор, Семиганч, Огзи-Кичик. Группа выделилась по признаку «индекс фасетирования», имеющего разброс значений от 35.96 до 44.3. Для группы характерны средние значения индексов леваллуа и леваллуа типологического (табл.6)

Тип 4 представляет памятники со значением индекса фасетирования от 25.6 до 27. В группу вошли памятники Худжи и Тоссор. Группа отличается очень высокими показателями леваллуа типологического при наличии низких показателей леваллуа.

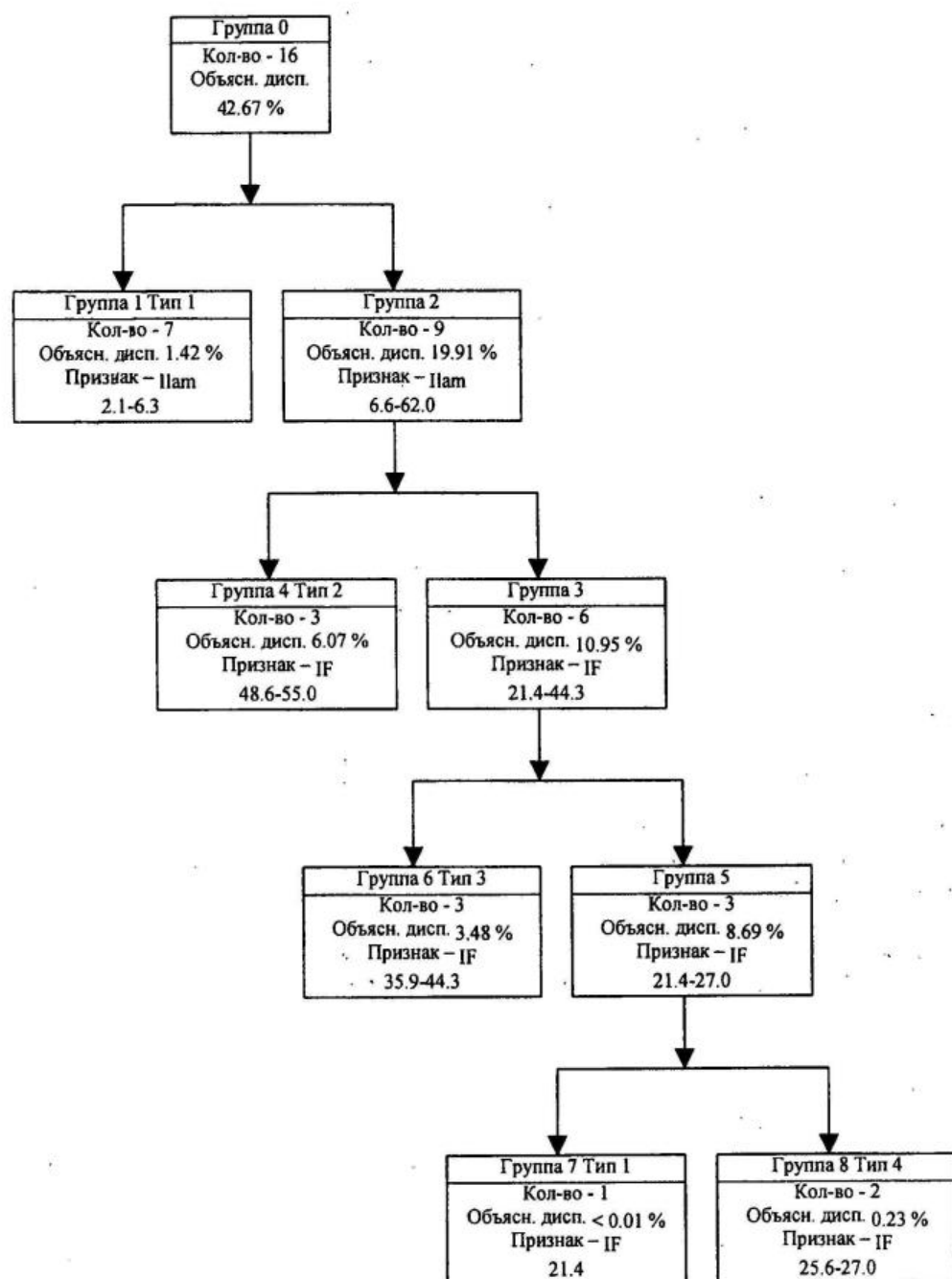


Рис. 2. Типологическое дерево.

Публикация результатов носит предварительных характер и необходимо проведение новых исследований для подтверждения того, что обнаруженная кластерная структура не является случайной флуктуацией. Для этого требуется выйти за пределы того признакового пространства, в котором была проведена классификация. Одним из способов убедиться в не случайности найденной кластерной структуры является сравнение классификаций, построенных на разных признаковых пространствах.

Таблица 6. Статистические характеристики сопряженности индексов и типов.

Типы	Число объектов	IL		ILty	
		Среднее	Станд. отклонение	Среднее	Станд. отклонение
1	8	6.31	3.49	19.52	9.02
2	3	40.97	2.53	56.77	25.61
3	3	20.80	9.48	28.60	1.70
4	2	13.55	1.15	82.50	4.50
Всего	16	16.41	13.90	36.08	25.77

Литература

- Артюхова О.А. Корреляция мустьерских индустрий Казахстана // Каменный век Казахстана и сопредельных территорий (материалы международной конференции, посвященной 70-летию Х.А.Алпысбаева). - Туркестан, 1998а:31-57.
- Артюхова О.А. Палеолитическое местонахождение Чингиз-I (Центральный Казахстан) // Каменный век Казахстана и сопредельных территорий (материалы международной конференции, посвященной 70-летию Х.А.Алпысбаева). - Туркестан, 1998б:58-69.
- Артюхова О.А. Мустье Казахстана // Хроностратиграфия палеолита Северной, Центральной и Восточной Азии и Америки. - Новосибирск. 1990: 35-39.
- Вишняцкий Л.Б. Палеолит Средней Азии и Казахстана. - СПб, 1996.
- Вишняцкий Л.Б. Изучение палеолита Средней Азии (материалы, методы, концепции). - СА, №1, 1989.
- Воронин Ю.А. Теория классифицирования и ее приложения. - Новосибирск, 1985.
- Гвоздовер М.Д., Григорьев Г.П. О факультетности в верхнем палеолите (по материалам Каменной Балки II). // КСИА. - М., 1975, вып. 141, с. 12.
- Гинзбург Э.Х., Горенштейн Н.М., Ранов В.А. Статистико-математическая обработка шести мустьерских памятников Средней Азии. // Палеолит Средней и Восточной Азии. - Новосибирск, 1980: 7-31.
- Гладилин В.Н. Проблема раннего палеолита Восточной Европы. - Киев, 1989: 230 с.
- Григорьев Г.П. Проблемы леваллуа. // МИА, № 185 (Палеолит и неолит СССР, т. VII. - Л., 1972: 68-74.
- Григорьев Г.П., Ранов В.А. О характере палеолита Средней Азии. // Тезисы докладов сессии, посвященных итогам полевых археологических исследований 1972 года в СССР. - Ташкент: ФАН, 1973.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Методы информатики в археологии каменного века. - Новосибирск, 1989.
- Деревянко А.П., Фелингер А.Ф., Холюшкин Ю.П. Опыт статистического группирования раннепалеолитических комплексов Евразии и Африки // Комплексные исследования палеолитических объектов бассейна р. Ануй. - Новосибирск, 1990:165-188.
- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т. Предварительные результаты информационно-статистического анализа мустьерских индустрий Алтая. // Методология и методика археологических реконструкций. - Новосибирск, 1994.
- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Ростовцев П.С. Некоторые статистические подходы к оценке факультетности позднего палеолита Енисея. // Методология и методика археологических реконструкций. - Новосибирск, 1994.
- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Ростовцев П.С. и др. Математические методы в археологических реконструкциях. - Новосибирск, 1995.
- Деревянко А.П., Холюшкин Ю.П., Ростовцев П.С. Неандертальская проблема как задача статистического анализа. // III Итоговая сессия Института археологии и этнографии СО РАН. Тезисы докладов. - Новосибирск, 1995: 47-49.
- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Ростовцев П.С. Некоторые статистические подходы к оценке факультетности мустьерских памятников Алтая. // Гуманитарные науки в Сибири, № 3, - Новосибирск, 1996: 3-10.
- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Ростовцев П.С. Структурный анализ мустьерских памятников Алтая. // Каменный век Казахстана и сопредельных территорий. - Туркестан, 1998: 93-111.

- Деревянко А.П., Холушкин Ю.П., Ростовцев П.С. Неандертальская проблема как задача статистического анализа (предварительные результаты)// Информационные технологии в гуманитарных исследованиях. - Новосибирск, 1998.
- Деревянко А.П., Холушкин Ю.П., Ростовцев П.С., Воронин В.Т. Статистический анализ позднепалеолитических комплексов Северной Азии. - Новосибирск, 1998.
- Деревянко А.П., Холушкин Ю.П., Ростовцев П.С., Воронин В.Т. Статистический анализ среднепалеолитических индустрий Ближнего и Среднего Востока. - Новосибирск, 1999.
- Дюран Б., Оделл П. Кластерный анализ. - М.: ИЛ, 1977.
- Жамбю М. Иерархический кластерный анализ и соответствия. - М.: Финансы и статистика, 1988.
- Замятин С.Н. О возникновении локальных различий в культуре палеолитического периода. // Происхождение человека и древнейшее расселение человечества. - М., 1951 (Труды Института этнографии, н.с., Т. 16).
- Исламов У.И., Оманжулов Т. Некоторые итоги изучения эпохи мустье Ташкентского оазиса. // ИМКУЗ - Ташкент, 1986 : 9 - 19.
- Касымов М.Р. Многослойная палеолитическая стоянка Кульбулак в Узбекистане (предварительные итоги исследования). // МИА, №185 (Палеолит и неолит СССР, т. VII. - Л., 1972: 111 - 119
- Коробков И.И. К проблеме изучения нижнепалеолитических поселений открытого типа с разрушенным культурным слоем. // Палеолит и неолит СССР. Т.6. - М.Л., 1971 (МИА, № 173).
- Коробков И.И. К вопросу о дивергентном характере эволюции древнепалеолитических индустрий (По материалам памятников Черноморского побережья Кавказа). // Actes du VIIe Congr. Intern. des sciences Préhist. Et Protohist. Prague (21-27 VIII 1966). - Prague, 1970.
- Коробков И.И. К вопросу о дивергентном характере эволюции древнепалеолитических индустрий (По материалам памятников Черноморского побережья Кавказа). // Доклады и сообщения археологов СССР на VII Международном конгрессе доисториков и протоисториков. - М., 1966
- Кулаковская Л.В. Мустье Азии: взгляд из Европы. // Хроностратиграфия палеолита Северной, Центральной и Восточной Азии и Америки. - Новосибирск. 1990: 210-214.
- Лбов Г.С. Методы обработки разнотипных экспериментальных данных. - Новосибирск: Наука, 1981, 1983.
- Любин В.П. К вопросу о методике изучения нижнепалеолитических каменных орудий// Палеолит и неолит СССР, Т. 5. - М.-Л., 1965: 7-75.
- Миркин Б.Г. Анализ качественных данных и структур. - М., 1980.
- Окладников А.П. Палеолит и мезолит Средней Азии// Средняя Азия в эпоху камня и бронзы. - М.-Л., 1966.
- Окладников А.П., Ранов В.А. Каменный век. // История таджикского народа. Т. 1., М., 1963.
- Ранов В.А. О восточной границе мустьерской культуры // Хроностратиграфия палеолита Северной, Центральной и Восточной Азии и Америки. - Новосибирск, 1990: 262-268.
- Ранов В.А. Парадокс леваллуа. // Каменный век. Памятники. Методика. Проблемы. - Киев: Наукова Думка, 1989: 46-50.
- Ранов В.А. Каменный век Южного Таджикистана и Памира. Автореферат диссертации на соискание степени доктора ист. наук в форме доклада. - Новосибирск, 1988: 52 с.
- Ранов В.А. Семиганч - новое мустьерское местонахождение в Южном Таджикистане. // МИА, № 185 (Палеолит и неолит СССР, т. VII) - Л., 1972: 100 - 110.
- Ранов В.А. К изучению мустьерской культуры в Средней Азии. // Материалы и исследования по археологии СССР, № 173. - Л., 1971: 209-232.
- Ранов В.А. О возможности выделения локальных культур в палеолите Средней Азии. // Известия Отд. Общ. Наук Тадж.ССР, вып. 3 (53), 1968: 3 - 11.
- Ранов В.А. Каменный век Таджикистана. - Душанбе, 1965: 145 с.
- Ранов В.А., Амосова А.Г. Раскопки мустьерской стоянки Худжи в 1978 году. // Археологические работы в Таджикистане. Вып. 18. - Душанбе, 1984 : 11-47.
- Ранов В.А., Амосова А.Г. Раскопки пещерной стоянки Огзи-Кичик в 1977 году. // Археологические работы в Таджикистане. - Душанбе, 1977, вып. XVII: 7 - 33.
- Ранов В.А., Лаухин С.А. Новые раскопки стоянки Худжи, Таджикистан (предварительное сообщение). // Палеоэкология плейстоцена и культуры каменного века Северной Азии и сопредельных территорий. Т. 2. - Новосибирск, 1998а: 345-351.
- Ранов В.А., Лаухин С.А. Проблемы геохронологии лессового палеолита Южного Таджикистана. // Палеоэкология плейстоцена и культуры каменного века Северной Азии и сопредельных территорий. Т. 2. - Новосибирск, 1998б: 352-357.
- Ранов В.А., Несмеянов С.А. Палеолит и стратиграфия антропогена Средней Азии. - Душанбе, 1973.

- Ранов В.А., Шарапов Ш., Никонов А.И.** Фауна млекопитающих, археология и геология стоянки Отзи-Кичик (Южный Таджикистан) // Доклады АН Таджикской ССР. - Душанбе, 1973, т. XVI, № 7: 60 - 63.
- Ростовцев П.С.** Статистическое согласование мер связи в анализе социально-экономической информации. // Экономика и математические методы, 1991, т. 27, вып. 1: 150-156.
- Ростовцев П.С., Костин В.С.** Автоматизация типологического группирования. Препринт № 137. - Новосибирск, 1995
- Сулейманов Р.Х.** Статистическое изучение культуры грота Оби-Рахмат. - Ташкент, 1972.
- Фелингер А.Ф.** Статистические алгоритмы в социологических исследованиях. - Новосибирск, 1985.
- Фелингер А.Ф., Холюшкин Ю.П.** Опыт статистического группирования палеолитических памятников Алтая. // Историография и источники изучения исторического опыта освоения Сибири, вып. 1. Досоветский период. - Новосибирск, 1988: 6-8.
- Холюшкин Ю.П.** Кластерный анализ технологических индексов мустьерских комплексов Средней Азии. // Археология и этнография Сибири и Дальнего Востока. Тезисы докладов XXXVIII Региональной археолого-этнографической студенческой конференции, посвященной 90-летию академика А.П.Окладникова. - Улан-Удэ, 1998: 164-165.
- Холюшкин Ю.П.** Новые направления в изучении палеолита Северной Азии. Диссертация на соискание ученой степени доктора исторических наук в виде научного доклада. - Новосибирск, 1992.
- Холюшкин Ю.П., Жилицкая Г.Ю.** К методике восстановления данных при анализе технологических индексов мустьерских комплексов Средней Азии. // Информационные технологии в гуманитарных исследованиях. - Новосибирск, 1998:
- Холюшкин Ю.П., Ростовцев П.С.** О методике исследования устойчивости выявленных структур на материалах мустье Средней Азии. // Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий. Т.IV - Новосибирск, 1998а: 516-517.
- Холюшкин Ю.П., Ростовцев П.С.** Структурный анализ технологических индексов мустьерских комплексов Средней Азии. // Каменный век Казахстана и сопредельных территорий. - Туркестан, 1998б: 211-215.
- Холюшкин Ю.П., Ростовцев П.С.** Проблема статистического обоснования критериев выделения мустьерских фаций на территории Средней Азии. // Гуманитарные науки в Сибири, № 3, - Новосибирск, 1997: 3-6.
- Bordes F.** Le Paleolithique inferieur et moen de Jabrud (Syrie) et la question du pre-Aurignacien. // L'Anthropologie, 59 (5-6), 1955: 486-507.
- Bordes F.** Principes d'une methode d'etude des techniques de debitage et de typology du Paleolithique ancien et moyen // L'anthropologie, 1950, v. 54: 19-34.
- Kholushkin Yu.P., Rostovtsev P.S.** Problem of statistical Grounding of the Criteria for identification of the Mousterian facies in the Central Asia. // Pre'histoire Europe'enne? Vol. 11, 1977.
- Larichev V., Kholushkin Yu., Laricheva I.** Lower and Middle Paleolithic of Northern Asia: Achievement? Problems and Perspectives. // Journal of World Prehistory, vol. 1, N 4.- New-York and London, 1987: 415-464.
- Movius H.L.** Early man and Pleistocene stratigraphy in Southern and Eastern Asia. // Papers of the Peabody Museum of American Archaeology and Ethnology, 1944, vol. 19, N 3.
- Ranov V.A.** The palaeolithic industries of the Central Asia: A Revision // IX Congrès UISPP, colloque VII " Le paléolithique inférieur et moyen en Inde, an Asie Centrale, en Chine et dans le sud-est Aciatique" / - Nice, 1976: 91-128.
- Ranov V.A.** Zentralasien Forschungen und Vergleichenden Archaologie. Band 4. Neue Forschungen zyr Altenzeit, 1979.
- Ranov V.A., Davis R.C.** Toward a New Outline of the Soviet Central Asian Paleolithic. // Current Anthropology, vol. 20, № 2, 1979: 249 - 262.
- Simon J.** Resampling: The New Statistics, Resampling Stats. - Arlington, VA, 1993.
- Simon J.** What some puzzling problems teach about the theory of simulation and the use of resampling. // American Statistician, vol. 48, 1994: 290-293.

При проведении автоматической классификации часто возникает вопрос о том, насколько выделенные программой классы отражают реальную структуру данных, а не случайную флуктуацию расположения точек в признаковом пространстве. В данной статье предлагается метод проверки неслучайности найденной кластерной структуры, основанный на сравнении классификаций, построенных на разных признаковых пространствах и возможно, разными методами.

Подобная задача возникла при структурном анализе среднепалеолитических индустрий Кавказа и Ближнего Востока [Деревянко, Холушкин, Ростовцев, Воронин, 2002]. Для сравнительного анализа классификаций были отобраны по критерию полноты данных 64 археологических комплекса.

Обратимся к постановке задачи. Для этого рассмотрим, как мы можем сравнить результаты двух классификаций. Таблицы сопряженности номеров кластеров двух классификаций вполне достаточно, чтобы оценить степень их согласованности:

Таблица 1. Таблица сопряженности результатов двух классификаций

Классификация 2	Классификация 1			Итого
	1	2	3	
1	3	3		6
2	4		19	23
3	5	6	8	19
4		13		13
5	3			3
Итого	15	22	27	64

Эта таблица построена по результатам двух классификаций, приведенным в таблице 2.

Наш план заключается в том, чтобы по таблице 1 определить:

- степень согласованности классификаций;
- статистическую значимость полученной величины путем построения функции ее распределения в условиях нулевой гипотезы.

Определение степени согласованности классификаций

Сформулируем требования, которым должен удовлетворять искомый показатель степени согласованности классификаций.

Во-первых, он должен быть нечувствителен к порядку нумерации классов. Это требование вытекает из того, что процедура автоматической классификации выделяет классы объектов, не учитывая их содержательной характеристики, а опираясь исключительно на особенности взаимного расположения объектов как точек в многомерном признаковом пространстве. Поэтому номер класса является не более, чем условным идентификатором.

Во-вторых, наш показатель должен измерять степень согласованности даже при несовпадении количества классов в сравниваемых классификациях, поскольку иначе его практическое применение будет неоправданно ограничено.

В-третьих, он должен давать максимальное значение (например, 1) при сравнении классификации с собой.

В качестве ближайшего аналога рассмотрим индикатор κ (каппа), впервые предложенный Дж.Козном в [Cohen J., 1960; Cohen J., 1968], и затем независимо – Г.Раушенбахом и А.Заславским [Раушенбах Г. В., Заславский А. А., с.126-141], и используемый для сравнения признаков, принимающих сопоставимые значения, например, результатов диагностики больных двумя врачами-экспертами.

* Работа выполнена при финансовой поддержке РФФИ (проекты № 01-07-90393, 03-06-80418).

Таблица 2. Исходные данные: результаты разбиения на кластеры 64-х памятников методами k-средних в пространстве S^2 (1) и иерархического кластерного анализа в пространстве S^3 (2)

№	Памятник	Классификация		№	Памятник	Классификация	
		№1	№2			№1	№2
1	Амуд В4	2	1	33	Ябруд 10	2	4
2	Амуд В2	2	1	34	Кударо I За	3	2
3	Кеу сл. I I	3	2	35	Кударо I Зб	3	3
4	Кеу сл. II	3	2	36	Кударо I Зв	2	3
5	Кеу сл III	3	2	37	Кударо I 4	3	2
6	Кеу сл V	3	2	38	Каркустакау	3	2
7	Кзар-Акил XXVIA	2	3	39	Тамарашени	1	2
8	Кзар-Акил XXVIB	3	3	40	Монашеская	3	2
9	Кзар-Акил XXVIA	3	3	41	Губский Навес	1	3
10	Кзар-Акил XXVIB	1	3	42	Малая Воронцовка	3	3
11	Кзар-Акил XXVIII	1	2	43	Таглар 2 сл.	3	2
12	Кзар-Акил XXVIII	3	2	44	Таглар 3 сл.	3	2
13	Кунджи	2	4	45	Таглар 4а	3	2
14	Варвази А	2	3	46	Таглар 4б	3	2
15	Варвази В	2	4	47	Таглар 5	3	2
16	Варвази С	2	4	48	Таглар 6	3	2
17	Варвази D	2	4	49	Ортвала-Клде I	1	3
18	Сефуним А	3	3	50	Ортвала-Клде II	3	3
19	Сефуним 12	2	4	51	Ортвала-Клде III	3	3
20	Сефуним 13	3	2	52	Ортвала-Клде IV	2	3
21	Сефуним VI	3	2	53	Ортвала-Клде V	3	3
22	Сефуним VII	1	2	54	Ортвала-Клде VI	2	3
23	Сефуним В	1	2	55	Ортвала-Клде VII	1	5
24	Сефуним С	3	2	56	Двойной Грот	1	3
25	Ябруд 2	2	4	57	Азых 3 сл	2	3
26	Ябруд 3	2	4	58	Среднехаджохская	3	2
27	Ябруд 4	2	4	59	Азых 6 сл	2	4
28	Ябруд 5	1	1	60	Медвежье	1	5
29	Ябруд 6	2	4	61	Лусакерт D	1	3
30	Ябруд 7	2	1	62	Лусакерт А	1	1
31	Ябруд 8	2	4	63	Газма	2	4
32	Ябруд 9	1	1	64	Баракаевская	1	5

Примечание. Пространство S^2 – двумерное признаковое пространство, полученное процедурой многомерного шкалирования путем проекции точек из многомерного исходного пространства признаков на двумерную плоскость при максимальном сохранении взаимных расстояний между ними. Пространство S^3 – такое же отображение в трехмерное пространство.

Для его вычисления применяется формула:

$$\kappa = \frac{p_d - p_e}{1 - p_e} \quad (1)$$

где p_d – сумма долей в диагональных клетках таблицы сопряженности;

p_e – сумма ожидаемых долей в тех же клетках в условиях независимости признаков.

Из приведенной формулы видно, что индикатор κ достигает максимального значения (равного единице), когда все недиагональные элементы равны нулю. Согласованность переменных считается слабой, когда значение κ не превышает 0,4, заметной или хорошей – при значениях 0,4–0,75, и высокой – при значениях более 0,75 [см. Landis, J.R. and Koch, G.G.; Флейс Дж., стр.233].

Позаимствуем отсюда идею оценки степени согласованности суммой долей в диагональных клетках. Поскольку в нашем случае однозначное соответствие номеров кластеров разных классификаций заранее не установлено, мы вправе сами установить это соответствие из каких-либо соображений.

Для начала рассмотрим квадратную матрицу, в которой количества строк и столбцов совпадают. Тогда установление соответствия сводится к перестановке строк и столбцов матрицы, после которой соответствующие друг другу строки и столбцы пересекаются на главной диагонали.

Когда же, как в рассматриваемом примере с трех- и пятикластерной классификациями, количества строк и столбцов различаются, лишние строки или столбцы присоединяем к соседним, то есть сопоставляем одному столбцу сразу несколько строк или одной строке несколько столбцов.

Устанавливать соответствие между номерами кластеров двух классификаций будем так, чтобы сумма частот на главной диагонали принимала максимально возможное для данной таблицы значение. Для нашей таблицы максимальная сумма 46 достигается при следующем порядке строк и столбцов:

Таблица 3. Таблица сопряженности после установления максимального соответствия

Классификация 2	Классификация 1			Итого
	1	2	3	
2	19		4	23
3	8	6	5	19
4		13		13
5			3	3
1		3	3	6
Итого	27	22	15	64

Делением этой величины на 64 получаем статистику согласованности:

$$p_d = 0,71875.$$

Распределение согласованности в условиях нулевой гипотезы

Сформулируем нулевую гипотезу. Поскольку предлагаемая статистика предназначена для измерения связи, то нулевая гипотеза должна предполагать отсутствие этой связи. То есть, при выполнении нулевой гипотезы результаты каждой классификации остаются теми же, но связь между ними разрушена. Условия нулевой гипотезы гарантированно выполняются в экспериментах с перемешиванием данных, когда значения первого признака остаются на своих местах, а значения второго перемешиваются случайным образом. Алгоритмически перемешивание реализуется случайной выборкой без возвращения. Проведя серию из 100 или 1000 таких вычислительных экспериментов, мы можем получить эмпирическое распределение, близкое к теоретическому. Чем больше будет проведено экспериментов, тем ближе полученное распределение к теоретическому.

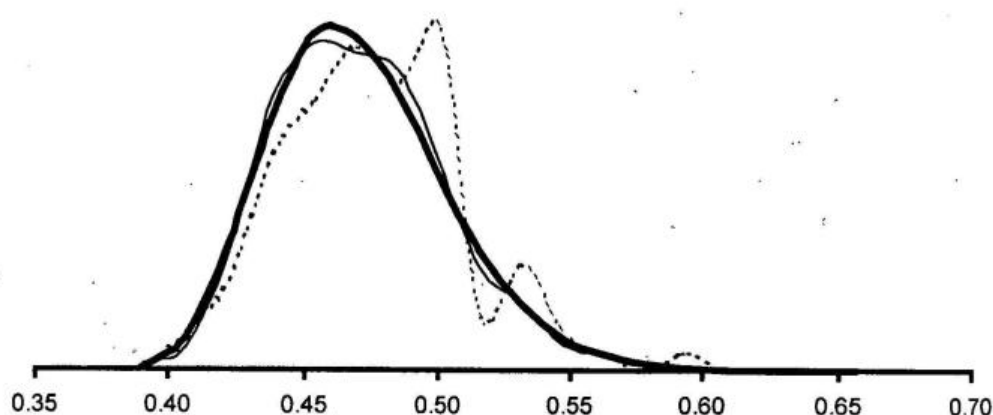


Рис. 1. Эмпирическая плотность распределения степени согласованности классификаций по результатам статистических экспериментов.

Условные обозначения: — 10^2 экспериментов
 - - - 10^3 экспериментов
 10^5 экспериментов

Однако, такой подход недостаточен для оценки вероятности редких событий. А как раз маловероятные в условиях нулевой гипотезы события и являются практически интересными. Здесь может быть предложено два варианта:

- аппроксимировать полученное эмпирическое распределение каким-либо известным параметрическим распределением, например, нормальным или бета-распределением;
- рассчитать точное теоретическое распределение.

Аппроксимация эмпирического распределения

Для проверки качества аппроксимации нормальным распределением мы провели 10 млн. статистических экспериментов с перемешиванием данных. В результате было получено распределение, приведенное в таблице 4.

Таблица 4. Эмпирическое распределение степени согласованности классификаций по результатам 10^7 экспериментов

Сумма на диагонали	Процент соответствия	Число экспериментов	Наблюдаемая частота	Значимость
25	39.1	17 325	0.0017325	1.0000000
26	40.6	176 180	0.0176180	0.9982675
27	42.2	709 246	0.0709246	0.9806495
28	43.8	1 415 180	0.1415180	0.9097249
29	45.3	1 874 771	0.1874771	0.7682069
30	46.9	1 867 066	0.1867066	0.5807298
31	48.4	1 543 552	0.1543552	0.3940232
32	50.0	1 072 880	0.1072880	0.2396680
33	51.6	656 910	0.0656910	0.1323800
34	53.1	359 611	0.0359611	0.0666890
35	54.7	176 942	0.0176942	0.0307279
36	56.3	79 562	0.0079562	0.0130337
37	57.8	32 484	0.0032484	0.0050775
38	59.4	12 221	0.0012221	0.0018291
39	60.9	4 172	0.0004172	0.0006070
40	62.5	1 356	0.0001356	0.0001898
41	64.1	390	0.0000390	0.0000542
42	65.6	120	0.0000120	0.0000152
43	67.2	29	0.0000029	0.0000032
44	68.8	3	0.0000003	0.0000003

Полученное эмпирическое распределение с хорошей точностью совпадает с теоретическим, которое будет рассчитано далее.

На рисунке 2 приведены результаты аппроксимации, из которых видно, что нормальное и бета-распределение дают близкие друг к другу функции плотности распределения, но оба не воспроизводят особенности наблюдаемого распределения. Следовательно, оценка значимости с использованием нормального или бета-приближения даст значительную ошибку при любом числе экспериментов, но может быть использована в качестве грубого приближения, если расчет теоретического распределения по каким-либо причинам невозможен.

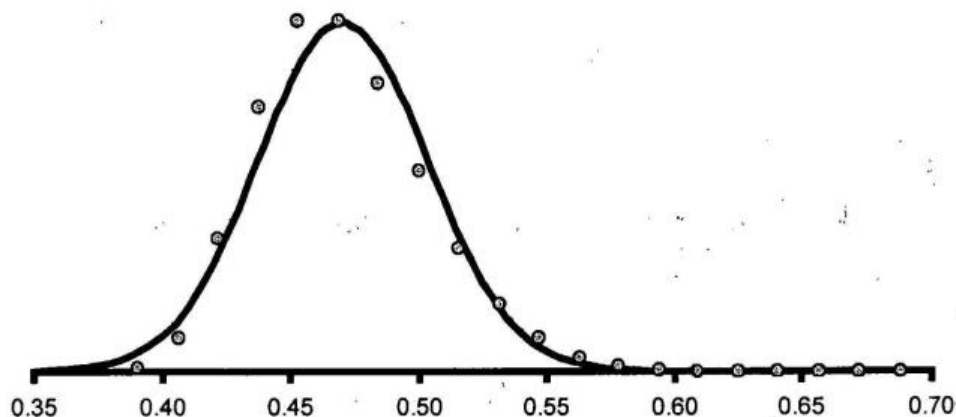


Рис. 2. Аппроксимация эмпирической плотности распределения.

Условные обозначения:

- эмпирическая плотность распределения по результатам 10^7 экспериментов
- аппроксимация Бета-распределением
- аппроксимация нормальным распределением

Для нашего примера получим:

$$P_{Norm}(L \geq 46) \approx 3.03 \cdot 10^{-14}$$

$$P_{Beta}(L \geq 46) \approx 2.26 \cdot 10^{-15}$$

Как мы увидим в дальнейшем, эти оценки значимости очень далеки от точного значения, полученного из теоретического распределения.

Нормальное распределение не является адекватной аппроксимацией, что подтверждается значимым отличием показателей асимметрии и эксцесса от нуля:

Таблица 5. Параметры аппроксимации эмпирической плотности распределения нормальным распределением

Число наблюдений	10 000 000
Среднее	0.470654 ± 0.000010
Стандартное отклонение	0.033045
Асимметрия	0.535087 ± 0.000775
Эксцесс	0.324155 ± 0.001549

Построение теоретического распределения

Построение точного теоретического распределения возможно, но требует больших затрат машинного времени на вычисления. Все же для рассматриваемого примера время вычислений оказалось приемлемым (порядка 5 минут) и теоретическое распределение было построено.

Опишем алгоритма расчета. Он распадается на решение следующих задач:

- расчет вероятности каждого варианта заполнения;
- полный перебор вариантов заполнения таблицы сопряженности с накоплением суммарной вероятности по значениям статистики согласованности.

Расчет числа элементарных событий для варианта заполнения.

Число элементарных событий, соответствующих каждому варианту заполнения таблицы, определяет вероятность его реализации.

Подсчитаем общее количество элементарных событий при перемешивании данных. Поскольку первый элемент может быть выбран одним из N способов, второй – одним из $(N-1)$ способов и т.д., то общее количество получается равным $N!$, или хорошо известному в комбинаторике числу перестановок для N элементов.

Для наших данных:

$$N! = 64! \approx 1.27 \cdot 10^{89}$$

Далее для примера подсчитаем, сколькими способами мы можем разместить 64 шара 5 цветов по 3 ящикам так, чтобы в каждом ящике было заданное число шаров каждого цвета.

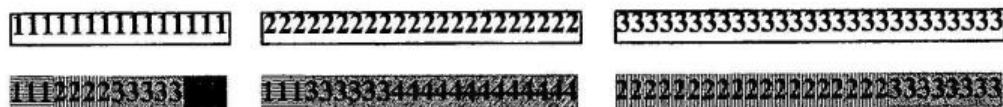


Рис. 3. Размещение шаров по ящикам. Количество шаров соответствует наблюдаемым частотам из табл. 2

Условные обозначения:

111, 222, 333 – ящики №1, 2, 3 соответственно;

11, 22, 33, 12, 13, 23 – шары пяти различных цветов.

Количество вариантов K , которыми можно разложить наши объекты по ящикам, вычисляется следующим образом:

$$K = (15! \cdot 22! \cdot 27!) \cdot \frac{6!}{3! \cdot 3! \cdot 0!} \cdot \frac{23!}{4! \cdot 0! \cdot 19!} \cdot \frac{19!}{5! \cdot 6! \cdot 8!} \cdot \frac{13!}{0! \cdot 13! \cdot 0!} \cdot \frac{3!}{3! \cdot 0! \cdot 0!} \quad (2)$$

Здесь первая скобка $(15! \cdot 22! \cdot 27!)$ содержит произведение числа перестановок внутри каждого ящика, то есть полное число комбинаций для случая, когда каждый объект жестко привязан к своему ящику. В следующих пяти дробях вычисляется количество вариантов распределения объектов каждого из пяти сортов между тремя ящиками без учета порядка расположения объектов внутри ящиков, который уже учтен в первой скобке. Обобщая это выражение, можем написать:

$$K = \frac{\prod_{i=1}^k N_{i.}! \cdot \prod_{j=1}^m N_{.j}!}{\prod_{i=1}^k \prod_{j=1}^m n_{ij}} \quad (3)$$

Для нашей таблицы:

$$K \approx 9.9 \cdot 10^{73}$$

Вероятность реализации определенного варианта заполнения таблицы можно получить, разделив K на полное число элементарных событий, равное $N!$:

$$P_K = \frac{\prod_{i=1}^k N_{i.}! \cdot \prod_{j=1}^m N_{.j}!}{\prod_{i=1}^k \prod_{j=1}^m n_{ij} \cdot N!} \quad (4)$$

Как можно заметить, мы получили гипергеометрическое распределение для таблицы произвольного размера. В нашем случае:

$$P_K \approx 7.8 \cdot 10^{-16}$$

Полный перебор вариантов заполнения таблицы сопряженности.

Перебор вариантов заполнения может быть реализован в виде рекурсивной процедуры, которая перебирает в цикле все варианты заполнения одной клетки таблицы и вызывает ту же процедуру для следующей клетки. Для определенности примем, что перебор начинается с левого верхнего угла таблицы, двигаясь вправо и вниз. Внутри процедуры требуется найти границы допустимых частот заполнения текущей клетки с учетом того, что частоты всех верхних клеток и левых клеток текущей строки уже заданы.

Рассмотрим клетку таблицы на пересечении строки i со столбцом j и найдем ограничения на частоту n_{ij} , накладываемые условиями постоянства итоговых частот. Для этого сгруппируем все строки и столбцы вне клетки ij .

Таблица 6. Параметры аппроксимации эмпирической плотности распределения нормальным распределением

Строки	Столбцы			Итого
	$\{1, i-1\}$	i	$\{i+1, m\}$	
$\{1, i-1\}$	$\sum_{p=1}^{i-1} \sum_{q=1}^{j-1} n_{pq}$	$\sum_{p=1}^{i-1} n_{pj}$	$\sum_{p=1}^{i-1} \sum_{q=j+1}^m n_{pq}$	$\sum_{p=1}^{i-1} N_{p\bullet}$
i	$\sum_{q=1}^{j-1} n_{iq}$	n_{ij}	$\sum_{q=j+1}^m n_{iq}$	$N_{i\bullet}$
$\{i+1, k\}$	$\sum_{p=i+1}^k \sum_{q=1}^{j-1} n_{pq}$	$\sum_{p=i+1}^k n_{pj}$	$\sum_{p=i+1}^k \sum_{q=j+1}^m n_{pq}$	$\sum_{p=i+1}^k N_{p\bullet}$
Итого	$\sum_{q=1}^{j-1} N_{\bullet q}$	$N_{\bullet j}$	$\sum_{q=j+1}^m N_{\bullet q}$	N

Чтобы найти ограничения, накладываемые на n_{ij} , запишем условия неотрицательности для каждой из четырех клеток, ограниченных двойной линией:

$$\left\{ \begin{array}{l} n_{ij} \geq 0 \\ \sum_{p=i+1}^k \sum_{q=j+1}^m n_{pq} \geq 0 \\ \sum_{q=j+1}^m n_{iq} \geq 0 \\ \sum_{p=i+1}^k n_{pj} \geq 0 \end{array} \right. \quad (5)$$

Теперь выразим неизвестные частоты через известные и n_{ij} :

$$\left\{ \begin{array}{l} n_{ij} \geq 0 \\ (N - \sum_{p=1}^{i-1} N_{p\bullet} - \sum_{q=1}^{j-1} N_{\bullet q} + \sum_{p=1}^{i-1} \sum_{q=1}^{j-1} n_{pq}) - (N_{i\bullet} - \sum_{q=1}^{j-1} n_{iq}) - (N_{\bullet j} - \sum_{p=1}^{i-1} n_{pj}) + n_{ij} \geq 0 \\ N_{i\bullet} - \sum_{q=1}^{j-1} n_{iq} - n_{ij} \geq 0 \\ N_{\bullet j} - \sum_{p=1}^{i-1} n_{pj} - n_{ij} \geq 0 \end{array} \right. \quad (6)$$

Отсюда получаем систему неравенств, определяющих границы изменения n_{ij} :

$$\left\{ \begin{array}{l} n_{ij} \geq 0 \\ n_{ij} \geq (N_i - \sum_{q=1}^{j-1} n_{iq}) + (N_j - \sum_{p=1}^{i-1} n_{pj}) - (N - \sum_{p=1}^{i-1} N_{p\bullet} - \sum_{q=1}^{j-1} N_{\bullet q} + \sum_{p=1}^{i-1} \sum_{q=1}^{j-1} n_{pq}) \\ n_{ij} \leq N_j - \sum_{p=1}^{i-1} n_{pj} \\ n_{ij} \leq N_i - \sum_{q=1}^{j-1} n_{iq} \end{array} \right. \quad (7)$$

Приведенных ограничений достаточно для реализации процедуры полного перебора частот заполнения в таблице сопряженности произвольного размера.

В результате расчета для нашей таблицы было получено следующее распределение:

Таблица 7. Распределение степени согласованности классификаций, полученное полным перебором вариантов заполнения таблицы сопряженности с использованием вероятностей (4) и ограничений (7)

Сумма на диагонали	Процент соответствия	Вероятность	Значимость	Вариантов заполнения
25	39.1	0.001749	1.000000	22
26	40.6	0.017560	0.998251	355
27	42.2	0.071013	0.980690	2258
28	43.8	0.141812	0.909677	8303
29	45.3	0.187544	0.767865	20943
30	46.9	0.186256	0.580321	41835
31	48.4	0.154406	0.394065	71618
32	50.0	0.107233	0.239659	107056
33	51.6	0.065733	0.132426	144617
34	53.1	0.035935	0.066693	182606
35	54.7	0.017707	0.030759	212644
36	56.3	0.007954	0.013052	237641
37	57.8	0.003262	0.005097	253386
38	59.4	0.001225	0.001835	258359
39	60.9	0.000423	0.000610	255557
40	62.5	0.000134	0.000187	242486
41	64.1	3.91E-05	5.31E-05	223009
42	65.6	1.05E-05	1.39E-05	198010
43	67.2	2.63E-06	3.39E-06	171144
44	68.8	6.06E-07	7.67E-07	143010
45	70.3	1.30E-07	1.61E-07	116589
46	71.9	2.57E-08	3.14E-08	91949
47	73.4	4.73E-09	5.69E-09	70683
48	75.0	8.10E-10	9.60E-10	52722
49	76.6	1.28E-10	1.50E-10	38563
50	78.1	1.88E-11	2.17E-11	27468
51	79.7	2.54E-12	2.89E-12	19250
52	81.3	3.14E-13	3.53E-13	13115
53	82.8	3.53E-14	3.93E-14	8734
54	84.4	3.60E-15	3.95E-15	5620
55	85.9	3.29E-16	3.57E-16	3541
56	87.5	2.66E-17	2.86E-17	2098
57	89.1	1.88E-18	2.00E-18	1219
58	90.6	1.14E-19	1.20E-19	623
59	92.2	5.72E-21	5.97E-21	293
60	93.8	2.39E-22	2.46E-22	108
61	95.3	6.77E-24	6.95E-24	32
62	96.9	1.79E-25	1.79E-25	8

Значимость нулевой гипотезы для нашего случая:

$$P(L \geq 46) = 3.14E-08$$

Исходя из "принципа практической невозможности маловероятных событий", гипотезу о независимости классификаций в нашем случае можно уверенно отвергнуть.

Подводя итоги, можно сказать, что предложенная нами статистика действительно позволяет проверять гипотезу о независимости результатов двух классификаций, построенных на одной и той же совокупности объектов. Эта статистика, в отличие от индикатора κ (каппа) обладает способностью сравнивать результаты классификаций без предварительного попарного отождествления кластеров одной классификации с кластерами другой. Несовпадение количества кластеров в обеих классификациях также не препятствует ее применению.

В ближайшем будущем предполагается применить предложенный в данной статье математический аппарат для анализа структур на многослойных памятниках Верхнего Енисея [Васильев, 2003].

Одним из возможных применений метода может быть получение обобщенной классификации по результатам сопоставления более чем двух исходных.

Литература

- Деревянко А.П., Холушкин Ю.П., Ростовцев П.С., Воронин В.Т. Корреляция среднепалеолитических индустрий Ближнего Востока и Кавказа. Новосибирск: Изд-во СО РАН, 2002. 186 с.
- Васильев С.А. Статистическое изучение динамики развития древней культуры Верхнего Енисея.
- Cohen J. A coefficient of agreement for nominal scales // Educ. Psychol. Measurement. 1960, v. 20, p. 37-46.
- Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit // Psychol. Bull. 1968, v. 70, p. 213-220.
- Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data. Biometrics, 33, 1977, p. 159-174.
- Раушенбах Г. В., Заславский А. А. Проверка однородности в задачах классификации: статистический подход и его табличное обеспечение // Материалы I Всесоюз. школы-семинара "Программно-алгоритмическое обеспечение анализа данных в медико-биологических исследованиях". Пушкино: НЦБИ АН СССР, 1986.
- Флейш Дж. Статистические методы для изучения таблиц и пропорций. М.: Финансы и статистика, 1989, с. 232-248.

Костин В.С.

Корнюхин Ю.Г.

Построение обобщенной классификации *

При структурном анализе среднепалеолитических индустрий Кавказа и Ближнего Востока [Деревянко, Холушкин, Ростовцев, Воронин, 2002] возникла задача: по результатам автоматической классификации 64-х археологических памятников, проведенной заранее разными методами (к-средних, иерархического кластерного анализа и типологического анализа) и на разных признаковых пространствах, построить некоторую сводную, обобщенную классификацию. Мы предлагаем метод решения этой задачи, основанный на анализе совпадения разных классификаций одних и тех же объектов. Рассмотрим его на примере десяти различных классификаций 64-х памятников.

Исходные данные для построения обобщенной классификации можно представить в виде таблицы объект-свойство, где объектами выступают памятники, а свойствами – номера кластеров, к которым они были отнесены в результате проведения каждой из классификационных процедур (см. табл. 2).

* Работа выполнена при финансовой поддержке РФФИ (проект № 03-06-80418).

Таблица 1. Характеристики классификаций

Номер классификации	Метод	Признаковое пространство	Число кластеров
1	k-средних	Факторы 1,2	5
2	k-средних	Факторы 1,3	4
3	k-средних	2-х мерные шкалы	3
4	k-средних	3-х мерные шкалы	5
5	Иерархический кластерный анализ	Факторы 1,2	5
6	Иерархический кластерный анализ	Факторы 1,3	4
7	Иерархический кластерный анализ	2-х мерные шкалы	3
8	Иерархический кластерный анализ	3-х мерные шкалы	5
9	Типологический анализ	% орудий от общего числа находок	5
10	Типологический анализ	% орудий от общего числа находок	3

Прежде, чем перейти к методу выделения наиболее устойчиво совместно классифицирующихся объектов в ядра кластеров обобщенной классификации, введем статистику для измерения степени близости объектов по результатам классификаций.

Измерение близости пары объектов.

Данные. Рассматриваются две реализации векторной случайной величины $C_i = \{c_i^1, c_i^2, \dots, c_i^m\}$ и $C_j = \{c_j^1, c_j^2, \dots, c_j^m\}$, где i и j – номера объектов в выборке, c_i^k – номер кластера по классификации k для i -го объекта, m – количество рассматриваемых классификаций. Размер всей выборки будем обозначать через N , а размер кластера c_i^k через $N_{c_i^k}$. Принадлежность i -го и j -го объектов к одному и тому же кластеру по классификации k обозначим через $c_{ij}^k = \delta(c_i^k, c_j^k)$, где $\delta(a, b)$ – дельта-функция Дирака, принимающая значение 0 или 1 в зависимости от совпадения или несовпадения аргументов, в нашем случае – номеров кластеров c_i^k и c_j^k :

$$\delta(c_i^k, c_j^k) = \begin{cases} 0, c_i^k \neq c_j^k \\ 1, c_i^k = c_j^k \end{cases}$$

Допущения. Классификация объекта j не зависит от классификации объекта i .

Нулевая гипотеза. Вероятность совпадения номера кластера c_j^k с номером кластера c_i^k равна вероятности для объекта j случайно занять одно из $N_{c_i^k} - 1$ мест соседей по кластеру объекта i :

$$H: P(c_j^k = c_i^k) = \frac{N_{c_i^k} - 1}{N - 1} \quad (1)$$

Поскольку эта вероятность не зависит от j -го объекта, будем обозначать ее как p_i^k .

Метод. Рассмотрим нормированные отклонения z_{ij}^k от ожидаемого значения:

$$z_{ij}^k = \frac{\delta(c_i^k, c_j^k) - p_i^k}{\sqrt{p_i^k(1 - p_i^k)}}$$

При совпадении и несовпадении номеров кластеров z_{ij}^k будет принимать значения:

$$z_{ij}^k = \begin{cases} -\sqrt{\frac{p_i^k}{1 - p_i^k}}, c_i^k \neq c_j^k \\ +\sqrt{\frac{1 - p_i^k}{p_i^k}}, c_i^k = c_j^k \end{cases}$$

Таблица 2. Результаты классификации памятников Кавказа и Ближнего Востока.

№	Памятник	Классификация									
		1	2	3	4	5	6	7	8	9	10
1	Амуд В4	3	1	2	4	1	1	1	1	1	1
2	Амуд В2	3	1	2	4	1	2	2	1	2	1
3	Кеу сл. I I	3	3	3	2	1	3	2	2	3	2
4	Кеу сл. II	3	3	3	2	1	3	2	2	3	3
5	Кеу сл III	3	3	3	2	1	3	2	2	3	1
6	Кеу сл V	3	3	3	2	1	3	2	2	3	1
7	Кзар-Акил XXVIA	5	3	2	3	2	4	2	3	4	1
8	Кзар-Акил XXVIB	5	3	3	2	1	3	2	3	3	1
9	Кзар-Акил XXVIIA	4	4	3	2	1	3	3	3	3	1
10	Кзар-Акил XXVIIБ	5	3	1	2	3	3	2	3	3	1
11	Кзар-Акил XXVIII	4	4	1	5	1	3	2	2	3	1
12	Кзар-Акил XXVIII	4	4	3	5	1	3	2	2	3	1
13	Кунджи	2	2	2	3	2	4	2	4	4	2
14	Варвази А	2	2	2	3	2	4	2	3	4	2
15	Варвази В	2	2	2	3	2	4	2	4	4	2
16	Варвази С	2	2	2	3	2	4	2	4	4	2
17	Варвази D	2	2	2	3	2	4	3	4	4	2
18	Сефуним А	5	3	3	2	1	3	2	3	3	1
19	Сефуним 12	2	2	2	3	2	4	2	4	4	1
20	Сефуним 13	3	4	3	2	1	3	2	2	3	1
21	Сефуним VI	4	4	3	5	1	3	2	2	3	1
22	Сефуним VII	4	4	1	5	1	3	2	2	3	1
23	Сефуним В	4	4	1	5	1	3	3	2	3	1
24	Сефуним С	3	4	3	5	1	3	3	2	3	1
25	Ябруд 2	5	3	2	3	2	4	2	4	4	2
26	Ябруд 3	5	3	2	4	3	2	2	4	5	2
27	Ябруд 4	2	2	2	3	2	4	1	4	4	2
28	Ябруд 5	1	1	1	5	4	2	3	1	4	2
29	Ябруд 6	1	3	2	3	4	3	3	4	4	2
30	Ябруд 7	1	1	2	4	3	2	3	1	3	2
31	Ябруд 8	5	3	2	3	3	2	1	4	5	2
32	Ябруд 9	1	1	1	1	4	2	3	1	4	2
33	Ябруд 10	5	3	2	3	3	3	1	4	4	2
34	Кударо I За	3	4	3	4	1	3	2	2	3	1
35	Кударо I 3б	4	4	3	4	1	3	2	3	3	1
36	Кударо I 3в	3	4	2	4	1	3	2	3	3	1
37	Кударо I 4	4	4	3	2	1	3	2	2	3	1
38	Каркустакау	4	4	3	2	1	3	2	2	3	1
39	Тамарашени	4	4	1	5	1	3	2	2	3	1
40	Монашеская	4	4	3	2	1	3	2	2	3	1
41	Губский Навес	5	3	1	1	3	3	3	3	4	1
42	Малая Воронцовка	5	3	3	2	3	3	2	3	3	2
43	Таглар 2 сл.	3	4	3	2	1	3	2	2	3	1
44	Таглар 3 сл.	3	4	3	2	1	3	2	2	3	3
45	Таглар 4а	3	4	3	2	1	3	2	2	3	1
46	Таглар 4б	3	4	3	2	1	3	2	2	3	1
47	Таглар 5	3	4	3	2	1	3	2	2	3	3
48	Таглар 6	3	3	3	2	1	3	2	2	3	3
49	Ортвала-Клде I	4	4	1	1	3	3	3	3	4	1
50	Ортвала-Клде II	5	3	3	2	1	3	2	3	3	1
51	Ортвала-Клде III	5	3	3	2	1	3	2	3	3	1
52	Ортвала-Клде IV	5	3	2	3	1	3	2	3	3	2
53	Ортвала-Клде V	5	3	3	2	1	3	2	3	3	1
54	Ортвала-Клде VI	5	3	2	3	1	3	2	3	3	1
55	Ортвала-Клде VII	4	4	1	1	3	3	3	5	5	1
56	Двойной Грот	1	4	1	1	4	3	3	3	4	2
57	Азых 3 сл.	3	3	2	4	1	2	2	3	5	1
58	Среднехаджохская	4	4	3	2	1	3	2	2	3	1
59	Азых 6 сл.	2	2	2	3	2	4	1	4	4	2
60	Медвежье	1	4	1	1	5	3	3	5	5	1
61	Лусакерт D	4	4	1	5	3	3	3	3	4	1
62	Лусакерт А	4	4	1	5	3	3	3	1	4	1
63	Газма	5	3	2	3	1	3	1	4	4	1
64	Баракаевская	1	3	1	1	4	3	3	5	5	2

Математическое ожидание z_{ij}^k равно нулю; а дисперсия – единице:

$$Mz_{ij}^k = \frac{0 - p_i^k}{\sqrt{p_i^k(1 - p_i^k)}} \cdot (1 - p_i^k) + \frac{1 - p_i^k}{\sqrt{p_i^k(1 - p_i^k)}} \cdot p_i^k = 0$$

$$Dz_{ij}^k = \left(\frac{0 - p_i^k}{\sqrt{p_i^k(1 - p_i^k)}} \right)^2 (1 - p_i^k) + \left(\frac{1 - p_i^k}{\sqrt{p_i^k(1 - p_i^k)}} \right)^2 p_i^k = \\ = \frac{(p_i^k)^2 (1 - p_i^k) + (1 - p_i^k)^2 p_i^k}{p_i^k(1 - p_i^k)} = p_i^k + (1 - p_i^k) = 1$$

Отклонение совпадений от ожиданий по m классификациям для i и j объектов обозначим через Z_{ij} :

$$Z_{ij} = \frac{1}{\sqrt{m}} \sum_{k=1}^m z_{ij}^k$$

Вычислим матожидание Z_{ij} :

$$MZ_{ij} = M\left(\frac{1}{\sqrt{m}} \sum_{k=1}^m z_{ij}^k\right) = \frac{1}{\sqrt{m}} \sum_{k=1}^m M(z_{ij}^k) = \frac{1}{\sqrt{m}} \sum_{k=1}^m 0 = 0$$

Аналогично вычисляется и дисперсия Z_{ij} :

$$DZ_{ij} = D\left(\frac{1}{\sqrt{m}} \sum_{k=1}^m z_{ij}^k\right) = \frac{1}{m} \sum_{k=1}^m D(z_{ij}^k) = \frac{1}{m} \sum_{k=1}^m 1 = \frac{1}{m} m = 1$$

Но дисперсия суммы равна сумме дисперсий только в случае, если z_{ij}^k для разных k независимы. Будем считать, что это условие выполнено, несмотря на то, что сами c_i^k для разных k не являются независимыми.

Далее заметим, что Z_{ij} изменяется в пределах:

$$Z_i^{\min} \leq Z_{ij} \leq Z_i^{\max}$$

где Z_i^{\min} определяется при несовпадении всех классификаций для i -го и j -го объектов:

$$Z_i^{\min} = -\frac{1}{\sqrt{m}} \sum_{k=1}^m \sqrt{\frac{p_i^k}{1 - p_i^k}},$$

а Z_i^{\max} - при полном совпадении:

$$Z_i^{\max} = \frac{1}{\sqrt{m}} \sum_{k=1}^m \sqrt{\frac{1 - p_i^k}{p_i^k}}$$

В частности, для всех диагональных элементов матрицы выполняется равенство:

$$Z_{ii} \equiv Z_i^{\max}$$

При выполнении нулевой гипотезы $H(1)$ эмпирическое распределение, построенное на наборе $\{Z_{ij}\}$ для всех $i \neq j$, не должно существенно отличаться от стандартного нормального распределения.

Мы промоделировали условия нулевой гипотезы на наших данных, проведя случайное перемешивание 64-х клеток внутри каждого из 10-ти столбцов таблицы 2. При этом были получены такие параметры:

Число наблюдений	4032
Среднее	0.000000±0.015518
Стандартное отклонение	0.985398
Асимметрия	0.249355±0.038561
Экссесс	-0.033223±0.077104

Видно, что все параметры, за исключением асимметрии, соответствуют стандартному нормальному распределению.

Это дает возможность проверить гипотезу о наличии кластерной структуры еще до выполнения процедуры классификации. Принятие нулевой гипотезы H будет означать, что все N объектов в среднем расклассифицированы по m классификациям независимо друг от друга и потому искомая обобщенная классификация будет не более, чем случайным результатом эвристической процедуры.

Если же гипотеза H отвергается, мы можем переходить к построению искомой обобщенной классификации. В нашем примере на исходных данных получены параметры распределения:

Число наблюдений	4032
Среднее	0.000000 ± 0.028853
Стандартное отклонение	1.832137
Асимметрия	0.606809 ± 0.038561
Экссесс	0.209070 ± 0.077104

Видно, что здесь стандартное отклонение достаточно далеко от единицы, что несовместимо с выполнением нулевой гипотезы.

Построение классификации.

Поскольку полученная статистика несимметрична, то есть $Z_{ij} \neq Z_{ji}$, эту меру нельзя использовать для проведения автоматической классификации стандартными методами. Поэтому переходим к описанию процедуры классификации, не требующей симметричности.

Данные. Имеется заполненная таблица $\{Z_{ij}\}$:

$$\{Z_{ij}\} = \begin{bmatrix} Z_{11} & \dots & Z_{1j} & \dots & Z_{1N} \\ \dots & \dots & \dots & \dots & \dots \\ Z_{i1} & \dots & Z_{ij} & \dots & Z_{iN} \\ \dots & \dots & \dots & \dots & \dots \\ Z_{N1} & \dots & Z_{Nj} & \dots & Z_{NN} \end{bmatrix} \quad (2)$$

Большие положительные значения Z_{ij} говорят о том, что объекты i и j оказываются в одном кластере существенно чаще ожидаемого в условиях нулевой гипотезы H .

Большие отрицательные значения Z_{ij} говорят о том, что объекты i и j оказываются в одном кластере существенно реже ожидаемого в условиях нулевой гипотезы H .

Кластерная структура. Кластер в матрице (2) представляется подмножеством строк (или столбцов с теми же номерами). Пересечение элементов этих строк и столбцов определяет блок матрицы. Удобно рассматривать матрицу после такой совместной перестановки строк и столбцов, которая собирает все клетки блока вместе. Тогда матрица приобретает блочно-диагональный вид. Внутри блоков значения Z_{ij} должны быть больше, чем за их пределами. В прямоугольных блоках на пересечении строк и столбцов из разных диагональных блоков, объединены показатели близости объектов, принадлежащих разным блокам, поэтому этот набор Z_{ij} характеризует степень контраста двух кластеров.

Критерий выделения кластеров. Зададим некоторое пороговое значение τ . Будем считать, что объекты образуют кластер только в том случае, когда гипотеза о равенстве нулю среднего значения Z_{ij} внутри блока отвергается на уровне значимости α . Это позволяет учесть не только среднее Z_{ij} , но и объем кластера. Проверку гипотезы можем проводить по Т-критерию Стьюдента.

Допущения. Выделенные блоки могут пересекаться. В этом случае кластером будем считать объединение блоков.

Некоторые объекты могут не входить ни в один из выделенных кластеров. Их будем считать выпавшими из обобщенной классификации.

Таблица 3. Выделение ядер кластеров обобщенной классификации

31	26	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100																	
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72																												

Метод. Предлагаемый метод относится к агломеративным методам кластерного анализа без обучения, использующих матрицу попарных взаимных близостей между объектами.


Опишем процедуру выделения кластеров.

На первом шаге переставим строки и столбцы матрицы, располагая вместе наиболее близкие объекты. С этой задачей успешно справляется программа поиска структуры в таблицах сопряженности [Ростовцев, Костин, Корнюхин, Смирнова, 1994: 60-61].

На втором шаге для всех возможных диагональных блоков вычисляем значимость гипотезы о равенстве нулю среднего значения Z_{ij} внутри блока. Если значимость меньше пороговой (), присоединяем блок к уже найденным.

Уменьшая пороговое значение, мы можем получить ряд кластерных разбиений, выделяющих все более и более устойчивые ядра. Таким путем можно построить своеобразную "карту уровней", напоминающую рельеф горной местности на физических картах.

Результат построения обобщенной классификации приведен в таблице 3, где диагональные блоки выделены оттенками серого в зависимости от пороговой значимости, на уровне которой среднее Z_{ij} в блоке отличается от нуля:

Пороговая значимость α	Цвет клетки
0.05	
0.01	
0.001	
0.0001	

Любой проведенный по границам области одного цвета блок представляет класс объектов, связанных сразу по нескольким исходным классификациям.

Таким образом, имея результаты нескольких заранее проведенных классификаций некоторой совокупности объектов, мы можем получить для них целый набор обобщенных классификаций: от мягких, охватывающих практически всю совокупность объектов и основанных на согласовании результатов всего лишь нескольких классификаций из исходного набора, до жестких, выделяющих лишь ядра кластеров из объектов, связанных наиболее значимо – сразу по множеству критериев, которые определяет исследователь выбором исходных классификаций.

Литература

- Деревянко А.П., Холушкин Ю.П., Ростовцев П.С., Воронин В.Т. Корреляция среднепалеолитических индустрий Ближнего Востока и Кавказа. Новосибирск: Изд-во СО РАН, 2002. 186 с.
- Ростовцев П.С., Костин В.С., Корнюхин Ю.Г., Смирнова Н.Ю. Алгоритмы анализа структуры и типологического группирования в анализе археологических данных// Методология и методика археологических реконструкций: Сб. науч. тр. - Новосибирск: Институт Археологии и Этнографии СО РАН, 1994. - С. 59-68.

Холюшкин Ю.П., РЕЛИГИОЗНО-МИФОЛОГИЧЕСКИЕ
Гемуев И.Н., ПРЕДСТАВЛЕНИЯ НАРОДОВ ЗАПАДНОЙ
Бауло А.В., СИБИРИ
Воронин В.Т.,
Нуртдинов А.Н.,
Ильиных М.Ю.

В секторе археологической теории и информатики (САТИ) ИАЭТ СО РАН с момента его образования в 1996 г. осуществляется комплексная программа исследований по созданию и развитию проблемно-ориентированной среды по гуманитарным наукам и разработке на этой основе гуманитарных информационных ресурсов.

Наиболее значимые элементы этой среды привязаны к археологическим и этнографическим исследованиям и представлению их результатов в Интернет [Деревянко, Холюшкин, Воронин и др., 2002: 6-14].

Важным направлением этих исследований и разработок является создание геоинформационной системы по религиозно-мифологическим представлениям народов Западной Сибири на основе материалов опубликованных монографий, статей, музейных коллекций Института археологии и этнографии СО РАН. Проектом создания этой системы предусмотрена организация свободного доступа пользователей по WWW к базе данных ГИС, исполнение этих запросов системой, разработка и встраивание прикладных программ для облегчения наполнения и редактирования базы данных. Создаваемая ГИС предусматривает использование в решении широкого круга задач этнографии, связанных с вовлечением в исследовательский процесс большого объема накопленных этнографических данных по религиозно-мифологическим представлениям народов Западной Сибири.

В рамках ГИС эти данные представлены в различных формах: текстовые материалы по истории исследования, фотографии и рисунки, привязанные к определенным ареалам. Система ориентирована на стандартные приемы и средства разработки ГИС, представление в Интернет, источники данных в виде опубликованных материалов этнографических исследований и наполнение данными результатов их обработки. Разработка рассчитана на широкий круг пользователей: этнографов, археологов, историков, научных работников, студентов, аспирантов, преподавателей вузов и школ.

Актуальность проекта этнографической ГИС обусловлена тем, что:

1) ГИС-системы в этнографии являются мощным комплексным информационным ресурсом, объединяющим разнородные этнографические данные в наиболее естественной для пользователя форме представления;

2) в среде исследователей, преподавателей, студентов и всех интересующихся мифологическими и религиозными представлениями народов Западной Сибири быстро возрастают потребности в интерактивных комплексных формах представления данных (Интернет, сетевых ГИС, гипертекстовых баз данных);

3) в настоящее время все крупные этнографические, археологические и музейные центры, высшие учебные заведения и культурно-просветительские учреждения имеют собственные web-узлы, с помощью которых пользователи могут получать доступ к информации, размещенной в Интернет.

* Работа выполнена при финансовой поддержке РГНФ (проект № 01-01-12015в).

Однако для этих и других категорий пользователей Интернет реально пока еще мало доступных по WWW баз данных в виде полноценных ГИС-систем не только по этнографии, но и по другим гуманитарным направлениям.

Географическая информационная система по мифологическим и религиозным представлениям народов Западной Сибири размещена на web-сервере Apache сектора археологической теории и информатики ИАиЭт СО РАН на серверной платформе Intel Pentium III.

Источниками информации для их внесения в базу данных ГИС и размещения их описания на страницах сайта системы являются результаты этнографических исследований и разработок, представленные в монографиях, статьях, обзорах, каталогах и т.д., в том числе малодоступных изданиях XVIII-XIX вв. На картах эти материалы привязаны к выделенным ареалам форм религиозно-мифологических представлений.

Одна из важных задач проекта связана с выбором платформы, на которой можно разместить создаваемую ГИС. Изучение и освоение существующих продуктов показал, что с точки зрения финансовых, кадровых и технических возможностей, которыми располагают участники проекта, наиболее подходящей платформой для размещения создаваемой системы является картографический сервер MapServer 4.0. Первоначальные версии этого продукта разрабатывались в США, университет штата Миннесота – University of Minnesota (UMN). Дальнейшее совершенствование и поддержку данного продукта проводит открытое сообщество OpenSource. Благодаря этому MapServer имеет свободное распространение.

Главная, с точки зрения реализуемости проекта, часть MapServer – это CGI-приложение для публикации GIS и снимковых данных в Интернет (WWW). Для просмотра данных в этих форматах достаточно иметь браузер Internet Explorer. Пакет также содержит несколько отдельных программ для создания карт, масштабных линеек и легенд, которые доступны клиенту сервера. В пакете также предусмотрены процедуры визуализации атрибутивной информации, привязанной к векторным картам, на сервере ГИС.

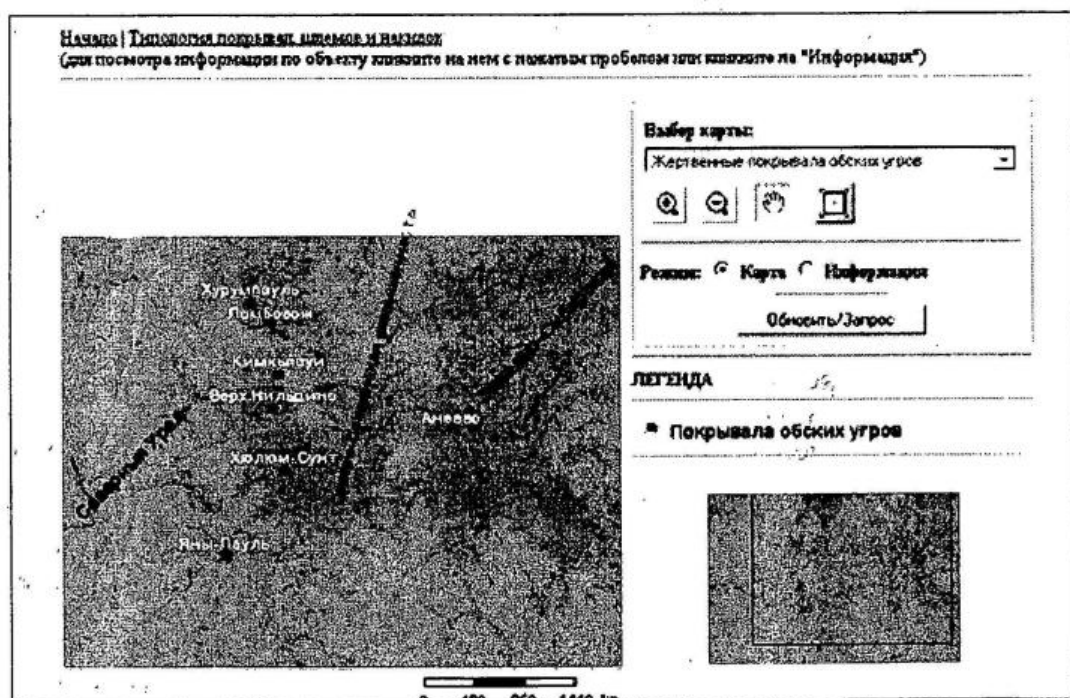


Рис. 1: Пример интерфейса ГИС (подсистема «Духовная культура»). Представлены жертвенные покрывала, шлемы, накидки и пояса обских угров.

Представлены также святилища Манси верховьев Сев. Сосьвы.

Для подготовки карт к размещению на сервере использовалась среда ArcView GIS Version 3.0 компании ESRI. Эта среда предоставляет возможность разработчикам

представить атрибутивную информацию в собственном формате (dbf-файлы) к shp-файлам. Одновременно информация геоинформационной системы представлена в формате XML, для работы с которой используется RHP-модуль.

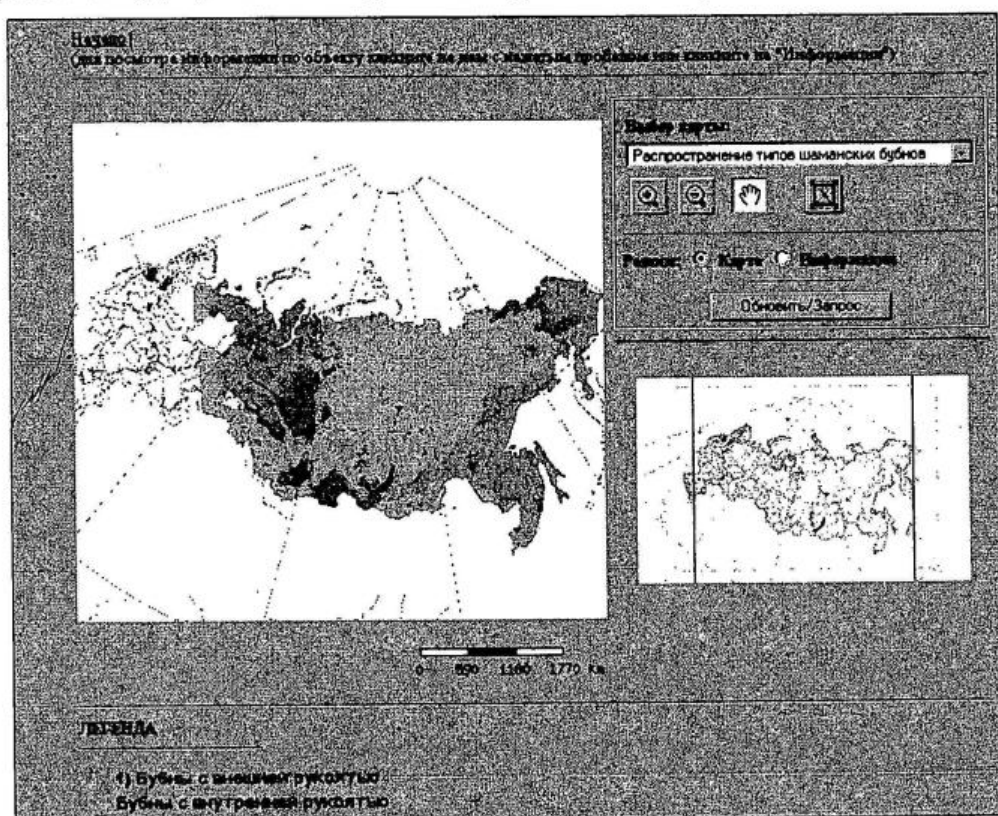


Рис 2. Пример интерфейса ГИС (подсистема «Материальная культура»). Представлены области распространения шаманских бубнов, типов жилищ и т.д.

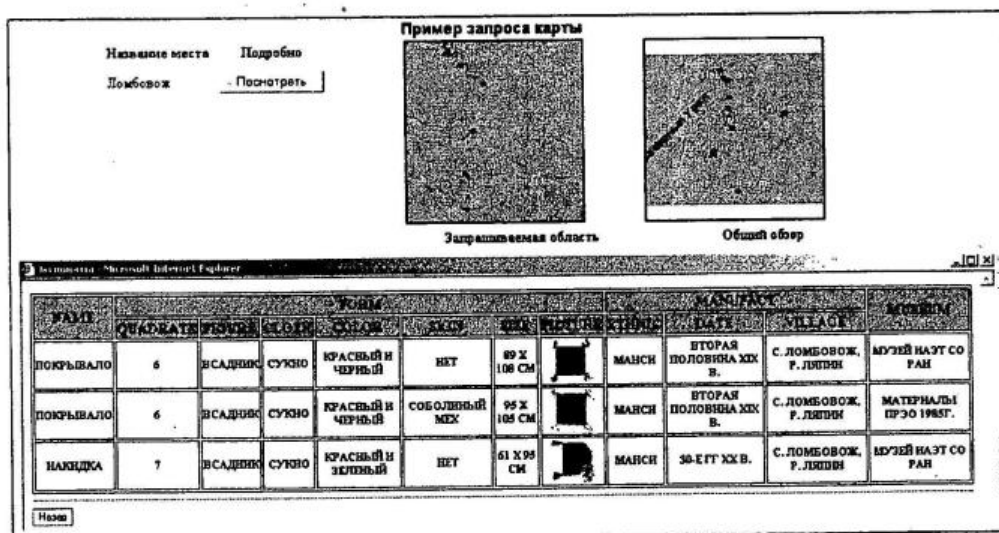


Рис. 3. Пример вывода атрибутивной информации. В верхнем окне представлены расположение запрашиваемого объекта и атрибутивная информация из БД формата dbf. С возможностью более подробного ознакомления (кнопка «Посмотреть»). В нижнем окне представлена атрибутивная информация их БД в формате XML.

Эти две базы данных (одна в формате dbf-файлов, другая – xml-формате) полностью независимы. Для связки этих баз данных в структуре записей dbf-файлов

предусмотрено некоторое поле. Оно позволяет привязать к каждому объекту на карте соответствующие позиции из базы данных в xml-формате.

Эта независимость баз данных и xml-форматы второй базы данных позволяют пополнять и редактировать наполнение последней в реальном времени с помощью php-скриптов.

Одна из проблем, возникших при создании нашей системы, связана с отсутствием векторных или растровых карт требуемой точности и объема. Для ее преодоления участники проекта применили две различные технологии размещения карт с содержательной информацией.

Первая технология была использована при разработке подсистемы "Духовная культура" (см. рис. 1, 3, 4). Здесь использовались снимки карт с CD-ROM "Атлас Мира; Atlas Deluxe of Rand McNally" с последующей обработкой в среде ArcView GIS Version 3.0 компании ESRI. С помощью этой технологии осуществлялись и добавление атрибутивной информации по соответствующим объектам, и собственно создание объектов. Таким образом, в результате комплексного применения подобной технологии на сервере размещались объекты, совмещенные с растровыми картами.

Вторая технология применялась при создании подсистемы "Материальная культура" (см. рис. 2). С ее помощью на сервере размещались растровые карты, отсканированные с «Историко-этнографического атласа Сибири», предварительно обработанные на Adobe Photoshop 6.0, с целью цветового выделения интересующих нас областей. В дальнейшем для привязки атрибутивной информации эти карты редактировались в ArcView 3.0.

При этом и в том, и в другом случаях объекты снабжались ссылками на имеющийся тематический материал.

Рис. 4. Пример интерфейса добавления новой записи в БД формата XML посредством php-скриптов (ГИС-подсистема «Духовная культура»).

Литература

- Деревянко А.П., Холюшкин Ю.П., Воронин В.Т., Воробьев В.В., Федоров С.А., Бердников Е.В., Елагина С.В. Создание информационного центра сектора археологической теории и информатики ИАЭТ СО РАН (проблемы и решения) // Информационные технологии в гуманитарных исследованиях. Вып. 3. Новосибирск: Редакционно-издательский отдел НГУ, 2002. С. 6-14.
- Derev'anko A.P., Kholyushkin Y.P., Voronin V.T., Ekimov D.V., Goriatchev D.N., Schipunov V.V., Kopteva N.V. Concepts of informational and statistical processing of archaeological data in the computer center of the Institute of Archaeology and Ethnography in Novosibirsk. // Computer Applications and Quantitative Methods in Archaeology 1994 (BAR International Series 600), 1995: 203-206.
- Гемуев И.Н., Люцидарская А.А. Служилые угры (один из аспектов русско-угорских отношений в XVI-XVII вв.) // Гуманитарные науки в Сибири, 1994 №3.

- Гемуев И.Н. Рец. на кн.: В.Е.Владыкин. Религиозно-мифологическая картина удмуртов. - Ижевск, 1994. - 383 с. // Гуманитарные науки в Сибири, 1996 №3.
- Гемуев И.Н., Пелих Г.И. О погребальной обрядности селькупов. // Acta Ethnographica Hungarica, 38(1-3), pp. 297-308 (1993).
- Историко-этнографический атлас Сибири. Под ред. М.Г.Левина и Л.В.Потапова. М.-Л.: Изд-во АА СССР, 1961. 498 с.

Холюшкин Ю.П.

Гемуев И.Н.
Воронин В.Т.
Фурсова Е.Ф.
Бауло А. В.
Воробьев В. В.
Грищенко А. А.

**Межэтнические взаимодействия и
межконфессиональное согласие
в Сибири (электронная библиотека как
инструмент решения проблемы)***

Актуальность проблемы межэтнического взаимодействия и межконфессионального согласия в Сибири в этнологии и этнографии обусловлена резко обострившимися в странах бывшего СССР и на планете в целом в течение последних десятилетий межнациональными и религиозными столкновениями и конфликтами и проявлениями религиозно-сектантского экстремизма, фанатизма и нетерпимости и соответственно угрозой распространения этих негативных явлений на сибирские территории, где они в условиях сосуществования полиэтничного и многоконфессионального населения Сибири пока менее выражены и латентны. Для поддержания межэтнических взаимодействий и установления межконфессионального согласия в Сибири требуются новые подходы и решения, связанные с воспитанием у населения, проживающего в зонах возможных и реальных конфликтов, мировоззрения, основанного на взаимоуважении и взаимопонимании

Современное состояние исследований по данной проблеме можно охарактеризовать как ситуацию социального заказа этнографической науке на выявление источников и причин резкого всплеска противоречий и конфликтов в межэтнических взаимодействиях и межконфессиональных столкновениях в мире в целом и в странах бывшего СССР, в частности, на поиск и отработку методов и инструментов для преодоления этих проблем. Глобальный характер проявления этих противоречий и конфликтов усиливает актуальность этого социального заказа. Как ответ на подобный социальный заказ этнографическая наука отреагировала ростом и расширением объема этнографических исследований и разработок по выявлению болевых точек и привлечению ним внимания властных структур. Однако результаты этих исследований и разработок пока еще не могут служить в полной мере инструментами решения указанных проблем. Исследовательских разработок по созданию специализированных в этом направлении библиотек, ориентированных на воспитание у населения ценностных установок взаимопонимания и взаимоуважения, не имеется.

Целью проекта является создание и поддержка электронной библиотеки по этнологии и этнографии многочисленных народов Сибири и Дальнего Востока и включение этих материалов в образовательно-просветительский процесс. Основу фонда электронной библиотеки составляют научные труды сотрудников Института археологии и этнографии (ИАЭТ) СО РАН. Накопленные за время существования Института и вошедшие в себя описания результатов сбора, обработки и научного

* Работа выполнена при финансовой поддержке РГНФ (проект № 00788atb) и Администрации Новосибирской области (договор № ФГ-10-03 от 31.03.2003 г.).

анализа богатейшего этнографического материала, эти публикации до сих пор малодоступны широкому кругу молодого и взрослого населения Сибири. В реализации проекта использованы и труды других исследователей культуры народов Сибири.

Имеющийся задел у коллектива, необходимый для успешной реализации этой цели и сложившийся на момент начала проекта, состоял из результатов выполнения участниками проекта аналогичных программ по созданию информационных систем в форме электронных каталогов и библиотек. Результаты этих раз разработок могут быть представлены в форме полнотекстовых баз данных, размещенных и доступных в Интернет: (http://sati.archaeology.nsc.ru/eas/index_r.html) – электронная полнотекстовая библиотека ИАЭТ СО РАН <http://www.sati.archaeology.nsc.ru/library>) – электронная полнотекстовая библиотека сектора этнографии народов Сибири ИАЭТ СО РАН. Обе библиотеки находятся в стадии разработки, но их фонды уже сейчас доступны в Интернет.

Главной задачей проекта является подбор и систематизация публикационного материала, анализ и выделение образовательного потенциала каждого издания и выбор адекватной формы его представления для свободного открытого доступа для населения Сибири. Создание подобной библиотеки и обеспечение широкого открытого и свободного доступа к ее фондам будет способствовать взаимопознанию этнических культур и препятствовать национальной розни и нетерпимости.

Научная новизна поставленной задачи состоит в том, что впервые в научной практике в области этнографии и этнологии в исследовательский процесс встраиваются не только сбор, обработка, научный анализ и представление результатов исследования в форме научных трудов, но и дополнительно последующий анализ и выделение образовательного потенциала и потенциала просвещения. Осуществляемые в духе культурного релятивизма в научных публикациях эти методы позволяют выявлять реальные возможности включения материалов научных исследований в образовательный и просветительский процесс. Новым направлением исследований является и представление этого потенциала в форме специализированной библиотеки с ориентацией на свободный открытый доступ для населения.

Для этих целей в проекте используются методы подбора, систематизации этнографического материала и экспертного анализа содержания научных публикаций, отражающих сведения о этнокультурных взаимодействиях и формах проявления межконфессионального согласия на территории Сибири. Отбор и анализ публикационного материала строится на понимании того, что этнокультурные и религиозные представления, составляющие концептуальную основу всех этносов и конфессий, опираются на общечеловеческие ценности (понятия добра, дружбы, ценности человеческой жизни, семьи, собственности, трудолюбия, благодарности, почитания старших, любви к родине и т.д.).

Наиболее важным моментом для проекта является положение о том, что для установления и поддержки взаимоуважения и взаимопонимания этносов и конфессий, построенных на понимании этнокультурных особенностей современных народов Сибири, необходимо знать историю их генезиса, эволюции и взаимодействия культур.

Происхождение того или иного народа, ныне живущего в исторически обозримое время, видится авторам проекта одной из наиболее значимых, а вместе с тем и сложных, трудно решаемых и дискуссионных проблем современной исторической науки. Исходя из этого, центральным объектом, на который должна быть нацелена просветительская, образовательная и воспитательная деятельность, является молодое поколение – будущее человечества. В этой связи важную научную проблему этой деятельности составляет выбор и обоснование формы, в которых она должна осуществляться, и ее информационной поддержки.

Важная роль отводится привлечению авторов научных публикаций к анализу образовательного потенциала их материалов для их последующего представления в образовательных ресурсах свободного открытого доступа. Это позволило образовательные ресурсы представить в современной электронной форме, разместить на web-сервере Института археологии и этнографии СО РАН и обеспечить

возможность свободного открытого для населения Сибири Доспа в Интернет. Для удобства пользования этими этнографическими информационными ресурсами участниками проекта разработан встроенный развитый сервис, основанный на отработанных механизмах навигации, контекстного поиска и этнографического тезауруса.

Особое внимание в проекте придано привлечению к процессу создания этнографической библиотеки авторов научных трудов.

Это дало возможность:

- 1) оперативно решать проблему авторских прав в юридическом и этическом плане;
- 2) в максимальной форме учитывать квалификацию, опыт и пожелания авторов при создании образовательных ресурсов на основе их публикаций;
- 3) принимать им непосредственное участие в образовательном процессе.

В соответствии с заявкой на грант в рамках создания электронной библиотеки осуществлена подготовка и размещение на ее страницах периодического издания "Сибирский этнографический вестник".

При реализации проекта важное внимание уделено использованию разработанной авторами структуры организации электронных полнотекстовых библиотек и технологии ее реализации. Основным принципом технологии является абстрагирование исходных данных от способа их представления. Подобный подход получает всё большее распространение в форматах метаданных (воплощаясь, например, в XML) и позволяет структуре информации (данным) быть независимой от конечного устройства вывода.

Этот подход, в частности, позволяет материалы электронной библиотеки, хранящиеся на сервере, делать доступными по любым существующим протоколам, когда эти материалы представлены в любом существующем формате. В частности, уже отработаны технологии доступа по сетевым протоколам http, Z39.50 к данным, представленным в форматах, традиционных для современного конечного пользователя: html/xml, текстовый, PDF.

Разумеется, большинство пользователей предпочитает использовать http протокол. Поэтому для подобного протокола организован интерфейс, позволяющий работать с электронной библиотекой, настраивать способ отображения текста в броузере, производить поиск.

Важным разделом электронной библиотеки по этнографии народов Сибири является форум в Интернет. Посетители библиотеки будут иметь возможность обмениваться мнениями по основной проблеме проекта и обсуждать содержание публикационных материалов, размещенных в ее информационном пространстве.

В соответствии с целями и задачами проекта в части его содержательного наполнения участниками были на начальной стадии реализации были решены следующие задачи:

- 1) постановка задачи конкретизации круга вопросов проблемы межэтнического взаимодействия и межконфессионального согласия в Сибири;
- 2) предварительный экспертный анализ содержания фонда электронных научных публикаций сотрудников Института археологии и этнографии СО РАН и других исследователей Сибири по этнографии на предмет освещения выделенных вопросов поставленной проблемы для последующего анализа их образовательного потенциала и потенциала просвещения в духе культурного релятивизма и включения в начальный фонд этнографической библиотеки.

В соответствии с этими задачами пополнен фонд открытой библиотеки по этнографии народов Сибири печатными изданиями с предварительной экспертизой соответствия их содержания поставленной проблеме. Отобранным публикациям придана электронная форма, обеспечивающая возможность использования в качестве образовательных ресурсов. Кроме того, осуществлены отработка технологий доступа к информационным ресурсам электронной библиотеки и анализ ситуаций использования ее фондов населением региона как инструмента преодоления основной проблемы, на решение которой направлен проект.

Научно-практическими результатами выполнения проекта являются:

1) функционирующая структура описанной выше полнотекстовой электронной библиотеки;

1) постановка задачи конкретизации круга вопросов проблемы межэтнического взаимодействия и межконфессионального согласия в Сибири;

2) предварительный экспертный анализ содержания фонда электронных научных публикаций сотрудников Института археологии и этнографии СО РАН и других исследователей Сибири по этнографии на предмет освещения выделенных вопросов поставленной проблемы для последующего анализа их образовательного потенциала и потенциала просвещения в духе культурного релятивизма и включения в начальный фонд этнографической библиотеки;

3) информационное наполнение фондов библиотеки электронными изданиями, размещенными на одном из серверов ИАЭТ СО РАН;

4) анализ и обобщение научных и технических проблем и научно-технических решений, направленных на их преодоление, в форме научных публикаций и отчета.

На момент написания настоящей статьи на страницах сайта "Этнография Сибири" размещены следующие информационные ресурсы, подготовленные участниками в течение выполнения проекта:

1) полнотекстовая библиотека работ ученых ИАЭТ СО РАН по этнографии народов Сибири;

2) видеотека документальных этнографических фильмов;

3) электронные этнографические энциклопедии и словари.

Полнотекстовая электронная библиотека работ ученых ИАЭТ СО РАН по этнографии народов Сибири включает:

– 14 монографий;

– три сборника научных трудов;

– шесть выпусков электронного журнала "Сибирский этнографический вестник" (№№ 11(12)-16(17)). Ответственный редактор: д.и.н. И.Н.Гемуев. Члены редколлегии: д.и.н. Н.А.Алексеев, д.и.н. Ф.Ф.Болонев, д.и.н. Ю.П.Холушкин;

– пять авторефератов докторских и кандидатских диссертаций по этнографии Сибири.

– библиографический указатель Чиндиной Л.А. "Археолого-этнографические издания Западной Сибири 1991-1997 гг."

Видеотека документальных этнографических фильмов представлена 6 фильмами по 20 минут.

Электронные этнографические энциклопедии и словари помимо информационного ресурса "Археология и этнография Приобья", размещенного на сайте на первом этапе, представлены на страницах сайта пятью электронными изданиями.

Всего на сайте размещено 121 научных публикаций, в состав которых, кроме монографий, научно-популярных фильмов, библиографических справочников и энциклопедий, включены 108 научных статей. Кроме того, на сайте "Этнография Сибири" были размещены выпуски электронного журнала "Сибирский этнографический вестник", подготовленные за предшествующие годы (включают 92 статьи).

Эти объемы этнографического материала на сайте "Этнография Сибири" делают его крупнейшим Интернет ресурсом в России по этой тематике. Несмотря на то, что сайт пока еще мало известен, его ежедневно посещают от 525 до 600 уникальных читателей (посетителей сайта без учета числа их заходов на страницы сайта в течение суток). Это доказывает, что сайт уже выполняет свою главную просветительскую функцию.

Литература:

Гемуев И.Н., Бауло А.В. Небесный всадник. Жертвенные покрывала обских угров.

Новосибирск, 2001. Электронная версия:

<http://www.sati.archaeology.nsc.ru:8101/library/nebo/>

Энциклопедия уральских мифологий. Т. 2. Мифология манси. Новосибирск, 2001.

Электронная версия приведена в энциклопедии "Археология и этнография Приобья": http://www.sati.archaeology.nsc.ru:8101/encys_p/.

Фурсова Е.Ф. Традиционная одежда русских крестьян-старожилов верхнего Приобья (конец XIX - начало XX вв.). Новосибирск, 1997.

Холюшкин Ю.П., Воронин В.Т., Федоров С.А., Бердников Е.В., Жилицкая Г.Ю.

Электронный каталог научной библиотеки ИАЭТ СО РАН // Информационные технологии в гуманитарных исследованиях. Новосибирск: Изд-во НГУ, 2002. С. 15-20.

Холюшкин Ю.П.

Воронин В.Т.

Воробьев В. В.

Бердников Е.В. Электронный каталог научной библиотеки

Федоров С.А. института археологии и этнографии СО

Жилицкая Г.Ю. РАН (археология и этнография)*

Грищенко А. А.

Лужин А.В.

Для расширения доступа к накопленным информационным ресурсам, сосредоточенным в фондах научной библиотеки, в Информационном центре ИАЭТ СО РАН, организованном в секторе археологической теории и информатики (САТИ), выполняется комплекс проектов и программ по созданию электронной библиотеки, включающей не только электронные библиографические базы данных научной библиотеки ИАЭТ СО РАН, но и библиотечные ресурсы, разработанные сотрудниками САТИ: Web-энциклопедии и информационно-справочные системы пилотного типа, электронные издания. Эти электронные материалы составляют основное содержание наполнения электронной библиотеки ИАЭТ СО РАН и служат предметом каталогизации ее ресурсов в форме электронного каталога. В ходе выполнения проектов отрабатываются технологии удаленного открытого доступа к этим информационным ресурсам.

Важное значение имеет проект по развитию электронного каталога по археологии и этнографии, выполняемый при финансовой поддержке Российского гуманитарного научного фонда.

Целью проекта является разработка электронного каталога по археологии и этнографии Сибири по изданиям и рукописям, хранящимся в научной библиотеке Института археологии и этнографии СО РАН. Электронный каталог представляет собой информационную систему со свободным доступом к базам данных по сети Интернет для научных сотрудников, преподавателей, студентов и учащихся. В рамках проекта предусмотрены разработка и встраивание прикладных программ для облегчения наполнения и редактирования базы данных. Ядро электронного каталога составляют библиографические данные о материалах, содержащих результаты научных исследований сотрудников Института археологии и этнографии СО РАН, хранящихся в фондах его научной библиотеки в форме редких книг, научных монографий, сборников научных трудов, периодических изданий, материалов международных, национальных и региональных научных конференций, препринтов, авторефератов диссертаций, рукописей диссертаций, собраний текстов рукописей, собранных в археографических экспедициях, и т.п. Большая часть этих материалов представлена в единичном экземпляре и не имеет подлинников и копий за пределами библиотеки. Другая часть опубликована ограниченным тиражом и малодоступна вне стен Института.

* Работа выполнена при финансовой поддержке РГНФ (проект № 01-01-12013в).

Использование данных электронного каталога позволит решить проблему доступа к фондам научной библиотеки Института археологии и этнографии СО РАН по сетям межбиблиотечного обмена и абонемента.

Проект направлен на решение проблемы информатизации фундаментальных научных исследований по основным направлениям гуманитарных наук, проводимых в научно-исследовательских учреждениях гуманитарного профиля Сибирского отделения РАН, высших учебных заведениях и музейных центрах Сибирского региона, Российской Федерации и других стран.

Круг пользователей охватывает сотрудников научно-исследовательских учреждений и подразделений гуманитарного профиля Сибирского отделения Российской академии наук и студентов гуманитарных кафедр и отделений колледжей и университетов сибирского региона, а также школьных учителей и старшеклассников, имеющих доступ в Интернет. Кроме того, в орбиту пользователей включается контингент из других регионов России и за рубежом.

Электронный каталог по археологии и этнографии Сибири размещается на Web-сервере сектора археологической теории и информатики ИАЭТ СО РАН на серверной платформе Intel Pentium III. Источниками информации являются фонды научной библиотеки ИАЭТ СО РАН, библиографические справочники и реферативные издания.

Современное состояние в области электронных каталогов характеризуется стремительным ростом числа электронных библиотек в России и за рубежом. Однако электронные каталоги гуманитарного направления по предметным областям представлены мало (от 300 до 700 наименований изданий в каждом). Библиографические материалы по авторам и названиям статей в периодических изданиях и сборников научных трудов в создаваемых электронных библиотеках не имеются. Настоящий проект ориентирован преимущественно на создание электронного каталога по материалам периодических изданий сборников научных трудов. Особое внимание в проекте уделяется включению в электронные каталоги материалов электронных изданий.

Коллектив разработчиков уже с 1998 года на инициативных началах выполняет программу по созданию и поддержке библиографических баз данных, размещаемых на Web-сервере информационного центра Института археологии и этнографии СО РАН (<http://www.sati.archaeology.nsc.ru>). В рамках этой программы были созданы и находятся в опытной эксплуатации три информационные системы:

- 1) база данных электронного каталога новых поступлений для научной библиотеки ИАЭТ СО РАН;
- 2) библиографическая база данных по публикациям научных сотрудников ИАЭТ СО РАН;
- 3) информационная система "Исследователи проблем арктического Севера", включающая основные публикации исследователей, их научные интересы и результаты.

В процессе создания и поддержки Web-сервера и информационных систем отрабатывались технологии, необходимые для реализации настоящего проекта, в том числе технологии организации и поддержки свободного доступа по сети, создания динамически изменяемых HTML-страниц и проектирования библиотечных БД с удалённым доступом.

В ходе реализации проекта выполнена комплексная междисциплинарная научнотеническая программа. Осуществлены освоение и эксплуатация прикладных программ для наполнения пилотного варианта электронного каталога в базе данных научной библиотеки ИАЭТ СО РАН (археология и этнография). При создании пилотного варианта в качестве программного обеспечения СУБД базы данных библиографических описаний была использована свободно распространяемое программное обеспечение СУБД WinIsis. Для его освоения и последующего использования потребовалась локализация, так как кириллизированных версий этого программного продукта не существует. На основе разработанной участниками проекта

кириллизованной версии программного обеспечения разработан единый формат данных электронного каталога и отработаны конвертеры.

Последовательное применение формата RUSMARC при построении моделей данных электронного каталога позволяет представить все библиографические и иные описания в единой манере, что дает возможность строить соответствующие схемы конвертации данных из различных небольших по объему библиотечных информационных систем и сводить их в более крупные интегрированные библиотечные базы данных.

Доработан конвертер электронного каталога из формата базы данных под управлением СУБД WinIsis в формат RUSMARC. Разработано и протестировано программное обеспечение Z-сервера (на Web-узле сектора археологической теории и информатики Института археологии и этнографии СО РАН) для последующего представления в Интернет пилотного варианта электронного каталога. Разработано и протестировано программное обеспечение Z-клиента (на Web-узле сектора археологической теории и информатики Института археологии и этнографии СО РАН) для доступа из Интернет к базам данных электронного каталога на Z-сервере.

Разработан и протестирован Web-сайт с доступом к базам данных электронного каталога на Z-сервере сектора археологической теории и информатики Института археологии и этнографии СО РАН. Осуществлено опытное подключение электронного каталога научной библиотеки ИАЭТ СО РАН к распределенному электронному каталогу Новосибирской региональной корпоративной библиотечной системы.

При работе над проектом учитывалось, что структура таблиц и триггеров (условий и требований, которым должны удовлетворять данные полей таблицы) базы данных электронного каталога является неоднородной, так как уже созданный ранее прототип электронного каталога по археологии и этнографии на основе фондов научной библиотеки Института археологии и этнографии СО РАН представляет собой собрание каталогов, существенно различающихся:

- по форме представления информационных ресурсов (библиографических данных, электронных коллекций, электронных публикаций и т.д.),
- по технологическим схемам ввода, корректировки и хранения данных,
- по методам доступа к ним конечного пользователя,
- формам использования данных в исследовательской практике историков, археологов и этнографов, а также учащихся, преподавателей, музейных сотрудников и других категорий пользователей фондов электронного каталога.

При сведении этих разнородных данных существует проблема разработки форм и правил единого их представления и модернизации технологических схем их использования в организации и ведении библиотечной работы.

Учитывая предыдущий опыт разработки полнотекстовых баз данных был принято во внимание то обстоятельство, что в настоящее время сформировалась отчетливая тенденция интеграции научных, образовательных и общественных библиотек в России и ее регионах на основе использования Интернет-технологий, открытых международных и российских стандартов, регулирующих форматы представления данных и методы доступа к ним. Особое значение в процессах интеграции российских библиотек имеют международные и российские стандарты и соглашения по форматам протоколов Web и Z39.50 и формату представления данных RUSMARC.

Протокол Z39.50 определяет порядок взаимодействия клиента и сервера, процедуры поиска и извлечения информации из баз данных и форматы представления этой информации. Однако этот протокол не определяет форматы хранения данных в конкретных базах данных, способы их индексации и процедуры функционирования различных СУБД. Кроме того, протокол Z39.50 не определяет интерфейсы взаимодействия пользователя и клиента. В идеологии Z39.50 в рамках одной схемы все базы данных совершенно одинаковы, несмотря на их физические различия по используемой СУБД, полям и синтаксису запросов. В Z39.50 невозможно определить, под управлением какой СУБД хранятся извлекаемые данные (Жижимов О.Л. Введение в Z39.50. Новосибирск, 2000).

Последовательное применение формата RUSMARC при построении моделей данных электронного каталога позволяет представить все библиографические и иные описания в единой манере, что дает возможность строить соответствующие схемы конвертации данных из различных небольших по объему библиотечных информационных систем и сводить их в более крупные интегрированные библиотечные базы данных.

На основе требований формата RUSMARC к форме представления данных библиографических описаний вначале была разработана опытная информационная система, включающая интерфейсные формы для пользователя и прикладные модули для наполнения и редактирования данных, ориентированная также и на predetermined запросы к ее содержимому. Однако тестирование этой системы показало, что требования указанных выше стандартов и протоколов эта система поддерживает не в той мере, которые нужны для обеспечения возможности интеграции электронного каталога научной библиотеки ИАЭТ СО РАН с другими библиотеками Сибирского региона и России.

Так, большое разнообразие форм и видов изданий и соответствующих им вариантов библиографических описаний (монографий, сборников научных трудов, отдельных статей, многотомных изданий, репринтных изданий, периодических и продолжающихся изданий, препринтов, авторефератов, диссертаций, научных отчетов, произведений одного или нескольких авторов в одной обложке, различных форм электронных ресурсов, размещенных на магнитных носителях или в Интернет и многое другое) достаточно детально может фиксироваться соответствующими макетами записей в форматах представления RUSMARC. Структура этих макетов достаточно сложна (ввиду большого числа полей и подполей с возможностью их повторения) и требует значительных усилий от разработчиков при проектировании баз данных библиографических описаний. Для того, чтобы выбрать из разнообразия форм представления данных наиболее подходящие, требуется сначала разработать первоначальные варианты схем и моделей данных, опробовать их на реальных массивах данных и в последующем использовать опыт и наработки для разработки более перспективных решений.

Для реализации этой стратегии авторы проекта выбрали свободно распространяемую СУБД CDS/ISIS for Windows (WinIsis), ориентированную на создание и поддержку библиотеко-подобных информационных систем. Хотя на освоение WinIsis ушло немало времени, с ее помощью исполнителям проекта удалось осуществить разработку специального конвертера, преобразующего библиографические записи в WinIsis в формат RUSMARC.

Первоначальная версия этого конвертера была ориентирована на периодическую перегрузку наполнения электронного каталога в специально разработанную базу данных, которая затем подвергалась автономному процессу конвертации и размещения в Интернет. В настоящее время отрабатывается новая версия системы, в технологии которой реализуется конвертация в реальном режиме времени.

На основе указанных исследований и разработок были предложены новые программно-технические и технологические решения, существенно расширяющие рамки проекта и повышающие его научную и практическую значимость. На базе этих решений появилась возможность не только построить и отработать новые наиболее современные технологические схемы, но и приступить к практическому решению ряда проблем, связанных с интеграцией информационных ресурсов научной библиотеки Института археологии и этнографии СО РАН (конвертацией базы данных новых поступлений, индексацией и каталогизацией полнотекстовых электронных баз данных, каталогизацией музейных коллекций и т.п.). В частности, на основе этих решений был разработан и опробован пилотный (опытный) вариант электронного каталога. Пилотный вариант размещен на сайте научной библиотеки Института (см. п. 1.10 настоящего отчета).

Основные результаты, полученные участниками по программе проекта, нашли непосредственное применение в информационных ресурсах библиотеки Института

археологии и этнографии СО РАН, в архитектуре и технологиях электронных баз данных, размещенных на страницах ее сайта. Опыт и наработки могут быть применены в дальнейшей работе над проектом, а также в аналогичных масштабных гуманитарных исследованиях и разработках.

Литература:

- Джим Грей о сегодняшнем и завтрашнем дне баз данных. СУБД, №3, 1998: 20-21.
- Бернстайн Ф. Middleware: модель сервисов распределенной системы. СУБД, №2, 1997: 41-60.
- Баженов С.Р., Мазов Н.А., Малицкий Н.А., Баженов И.С. Создание программного комплекса доступа из Интернет к базам данных на основе WWW-ISIS // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Материалы конф. "Крым 98". – М., 1998. – Т. 2: 581-584.
- Баженов С.Р., Мазов Н.А., Малицкий Н.А., Баженов И.С. Создание программного комплекса доступа из Интернет к базам данных на основе WWW-ISIS// Научные и технические библиотеки.– 1999. N2: 47–52.
- Галатенко В., Таранов А. Компонентная объектная модель JavaBeans. СУБД, №4, 1997: 42-60.
- Жижимов О.Л., Мазов Н.А., Фролов А.С. Доступ к базам данных ISIS из Internet и построение распределенной информационной системы. Вычислительные технологии, т.2, №3, 1997: 45-50.
- Мазов Н.А., Малицкий Н.А., Баженов С.Р., Баженов И.С., Жижимов О.Л. WEB-ориентированная информационно-поисковая система для доступа к базам данных CDS/ISIS // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества: Материалы 6 Междунар. конф. "Крым 99". – Судак, 1999. – Т. 1: 113 – 115.
- Племнек А.И., Усманов Р.Т. Z39.50: открытый доступ к библиографической информации. Научные и технические библиотеки, №8, 1998: 24-28.
- ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50-1995, July 1995.
- BIREME/PAHO/WHO. WWWISIS: a word-wide web server for ISIS-databases. Ver. 3.0. Sao Paulo, Brazilian, Oct. 1997.
- BIREME/PAHO/WHO. ISIS Application Program Interface: ISIS_DLL User's Manual: Sao Paulo, Brazilian, Aug. 1997.
- Index Data. YAZ User's Guide and Reference. Version 1.4. (<http://www.indexdata.dk>).
- Index Data. Zebra Server – Administrators's Guide and Reference. Revision 1.38. (<http://www.indexdata.dk>).

Ларичев В.Е.

МЕДВЕДЬ В ИСКУССТВЕ РАННЕГО ЭТАПА ВЕРХНЕГО ПАЛЕОЛИТА СИБИРИ (семантика образа в свете календарно- астрономических записей)*

Вводные замечания. Открытие М.М. Герасимовым антропоморфного облика скульптур, а также иного вида предметов искусства малых форм, показало несостоятельность гипотезы отсутствия «художественной одаренности» у обитателей Северной Азии ледниковой эпохи [Петри, 1923; 1927; 1928; Герасимов, 1931 а и б]. Сомнения в подлинности находок (наветы о подделках) были вскоре рассеяны, а идеологический донос (30-е годы!) и археологический разнос не имели, к счастью, последствий [Окладников, 1932]; подробности см. [Ларичев, 1998; 1999, стр. 147–154]. Удачи раскопок последующих лет превратили искусство Мальты в явление феноменальное для всей Азии. Таковым оно остается доселе, возбуждая, однако, споры об истоках его.

В 1976 г. начались исследования верхнепалеолитического поселения Малая Сья, которое значительно (\approx на 10 000 лет) превосходило по возрасту Мальту (34500 ± 450 ; 33060 ± 300) [Ларичев, 1978]. Памятник этот был уникальным для 70-х годов XX в. не только вследствие неожиданно глубокой для Сибири древности его в рамках хронологии верхнего палеолита, что позволило рассматривать в новом свете проблему становления (истоков) культур раннего *Homo sapiens sapiens* в пределах Северной Азии и Дальнего Востока (сообщение о том было сделано мною в том же, 1976 г. на Всемирном конгрессе археологов в Ницце). Уникальность его в той же мере определялась неожиданностью открытия в первый же год раскопок объектов искусства малых форм (см. рис. 1, а–в, е) [Ларичев, 1976; 1978 а и б; 1980 а и б]. Они, эти объекты, открывали простор для решения второй, не менее сложной и существенной проблемы сибирского палеолитоведения – *отыскания местных корней мальтинского искусства*, ибо стало ясно, что художественное творчество зародилось на Севере Азии примерно в то же время, что и в Европе. Сама же специфика образцов малосыйского искусства (среди них были, в частности, объекты с календарно-астрономическими записями [Ларичев, 1984] и каменные навершие bâton, подобное мальтинскому [Ларичев, 1992]) позволила выдвинуть гипотезу о космогонии и космологии в мировоззрении раннего *Homo sapiens sapiens* Сибири, а также «космичности» мышления его.

Все это вместе взятое и кое-что побочное, не связанное с наукой (почему о том нет смысла говорить), предопределило драму Малой Сьи, которой по удачному стечению обстоятельств удалось избежать Мальте (работу следственной «Комиссии» по материалам раскопок М.М. Герасимова в 1929 г. направлял человек чести и совести – Г.Ф. Дебец). Случилось то, что В.И. Матюшенко деликатно оценил в историографии археологии Сибири событием, «которое не красит науку» [Матюшенко, 2001]. А если назвать вещи своими именами, вне академических условностей, то тогда происшедшее четверть века назад *следует определить как расправу, исполненную в точном соответствии с иезуитски разбойными канонами печальной памяти 30-х годов*. Тут был полный набор обычных для того трагического времени аксессуаров, выстроенных в ряд знатоком дела:

* Исследование поддержано Российским фондом фундаментальных исследований (проект 02-06-80094) и Федеральной целевой программой «Интеграция».

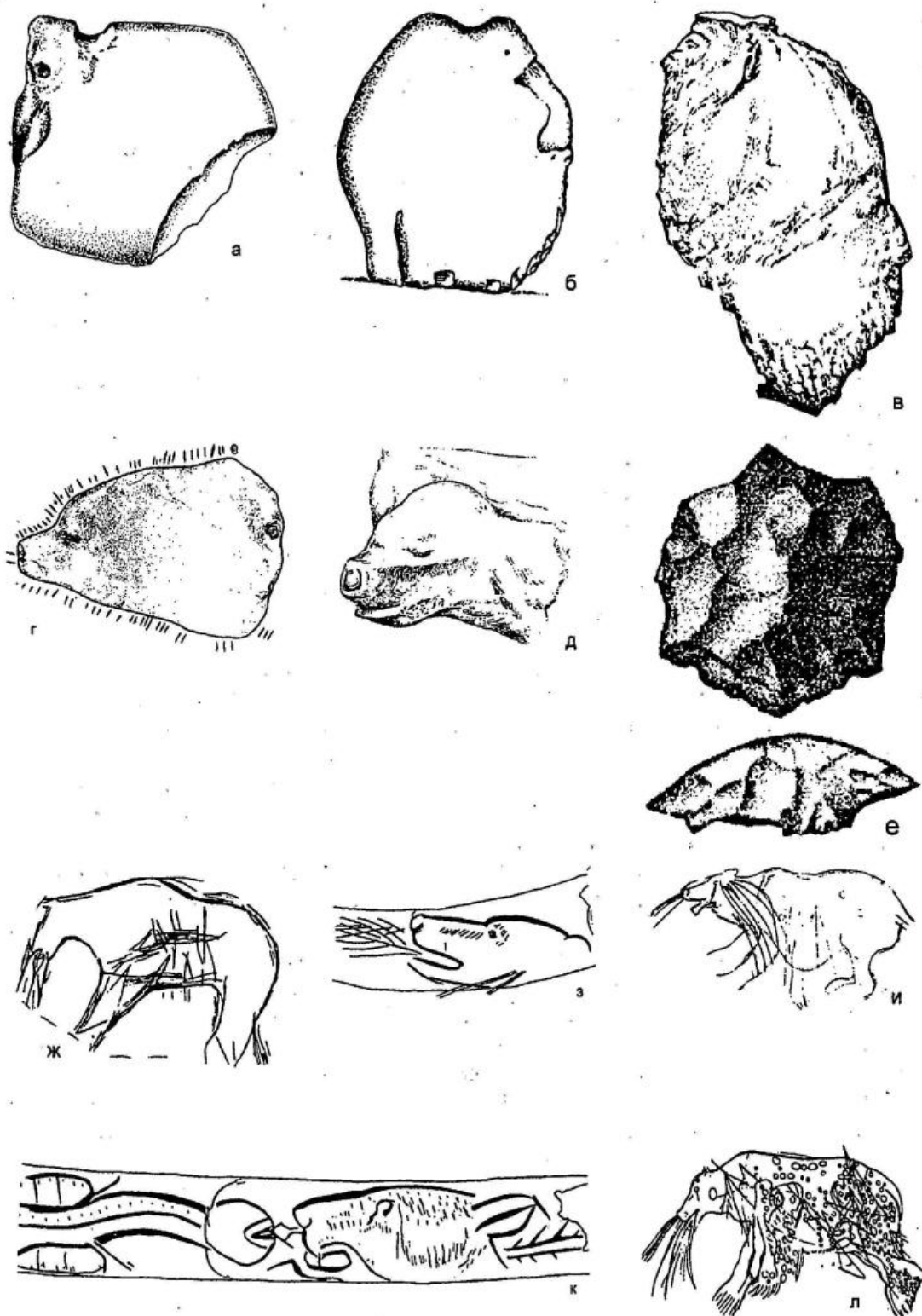


Рис. 1. Скульптуры Малой Сыи: а – мамонта; б – мамонта и бизона (стрелка указывает на голову и рог последнего); в – мамонта и бизона; е – черепахи; техника исполнения – резьба, пикетаж, скалывание, ретуш, окрашивание; скульптура Новой Сыды: г – схематический рисунок двуконечной скульптуры медведя (разворот влево) и мамонта (разворот вправо); краевые насечки и прочие знаки (всего их 69) вынесены за пределы контура скульптуры; резьба на широкой плоскости не зафиксирована; скульптура Толбаги: д – голова медведя; ж–л – медведь в искусстве палеолита Западной Европы.

и «Письмо в редакцию» с издевательскими оценками нечистой на руку личности – В.Е. Ларичева, идеологически чуждого и вредоносного советской науке (текст «Письма» был подготовлен в Новосибирске геологом Л.В. Фирсовым, переправлен в Ленинград, где его «подработал» А.Д. Столяр; он же собрал подписи у статистов-соучастников, отобранных А.П. Окладниковым из числа своих бывших сослуживцев (А.Н. Рогачев, М.П. Грязнов), и направил в «Советскую археологию» [см. № 4, 1981]);

и заключение «тройки» – следственной «Комиссии», негласно созданной «по просьбе академика А.П. Окладникова», спешно доставленной в Новосибирск для разбирательства деяний злостного и неумного распространителя ереси и оплаченной им [Любин, Абрамова, Борисковский, 1981, № 4] («Комиссия» констатировала, что никакого искусства в материалах Малой Сии нет);

и, наконец, резолюция самого организатора действия о «закрытии проблемы искусства Малой Сии, как ее трактовал В.Е. Ларичев» (некоторые подробности см. [Ларичев, 1999, стр. 154–156; 201–205; 331–335]).

В результате, изучение материалов Малой Сии было насильственно прервано на несколько лет.

Малая Сия (прежде всего – искусство этого памятника, а также идеи о содержательности его) стали жертвой специалистов – творцов теории «неудобных фактов», псевдорадетелей «чистоты науки», а также их подручных. Факты эти не устраивали их по разным причинам, одинаково некасающихся науки и потому не подлежащих обсуждению. Подоплека (мотивы) их действий тоже были разными, не подлежащими обсуждению по той же причине – нестыковке с целями науки. *Скрытый же мотив действий самого организатора фарса определялся, как и в случае с Мальтой, осознанием начала разрушения вековой давности парадигм палеолитоведения Сибири, в первую очередь – искусствоведения древнекаменного века, а с ним и традиционной концепции восприятия древнего человека существом примитивным, лишенным высокого статуса интеллекта и духовности.* Разрушения, к которому он, увы, (как и в болезненно досадном для него случае с Мальтой), не имел никакого отношения. А мог бы иметь, не отклони я дерзостно предложение патрона о соавторстве в публикации одного из предметов искусства Малой Сии – гравюры быка, исполненной на гальке.

Изучение предметов искусства Малой Сии было восстановлено лишь через несколько лет [Ларичев, 1984; 1987 а, 1992], а за сим последовали полтора десятилетия самого тщательного анализа предметов искусства Мальты под *астроархеологическим, палеоастрономическим и палеокалендарным* углами зрения. Эта исследовательская программа подтвердила правильность идей, которые были высказаны в первых статьях о содержательной сути объектов искусства Малой Сии. Что касается *времени зарождения художественного творчества в Сибири*, то вопрос этот решился сам собой, по ходу раскопок палеолитических памятников – в 1981 г. в печати появилось сообщение об открытии в Толбаге (Забайкалье, долина р. Хилон) скульптурного изображения головы медведя (см. рис. 1, д; дата памятника – синхронна Малой Сие – $34\,860 \pm 2\,100$ лет; $27\,210 \pm 30$ лет – малохетское (до 33 000 лет) или липовско-новоселовское (30–25 000 лет) потепление каргинского межледникового [Константинов и др., 1981; Константинов и др., 1983; Авдеев, 1986]), а в 2001 г. при обследовании местонахождения Новая Сыда, расположенного \approx в 100 км от Малой Сии, была найдена двуконечная, с числовыми знаковыми записями каменная (из стеатита) скульптура медведя и мамонта того же примерно возраста (рис. 1, з) [Ларичев, Аннинский, 2004, в печати]. Так разработка проблемы раннего этапа верхнего палеолита Сибири и *зарождения искусства в культурах его* обрели законную силу, а события 1929–1932 и 1979–1981 гг. навсегда ушли в историю археологии Сибири, став всего лишь хорошим поводом порассуждать, при случае, на излюбленную тему философов XVIII в. о «врожденном нравственном законе человека», то бишь, – о совести [Ларичев, 1999, стр. 144–154; 155–156; 202–205].

Постановка проблемы. Главные задачи искусствоведения палеолита России – избавление его от пустопорожних («философических») разговоров и выбор методики,

которая обеспечила бы *доказательность семантическим оценкам образов и композиций*, а с нею, доказательностью этой, достойное науки решение проблемы происхождения художественного творчества. Такой, пифагорейского стиля метод («Число объясняет Все»), был разработан в ходе изучения предметов искусства Мальты, Малой Сии, Ачинского поселения, Сунгири, Игнatieвской пещеры и bâtons de commandement, а также прочих объектов искусства западноевропейских памятников древнекаменного века, в том числе – пещерных (см., для примера [Ларичев, 1989; 1993; 2003]). Что касается образа медведя в искусстве палеолита Европы, то этому сюжету была посвящена специальная публикация (см. рис. 1, з, к [Ларичев, 2002]; там же – обоснование неприемлемости старых методических подходов при строгой нацеленности искусствоведа на интерпретации). Это и определило выбор темы публикации – семантика образа медведя в искусстве палеолита Сибири. Она, тема эта, призвана, помимо прочего, наглядно продемонстрировать – какова была реальная значимость всего лишь одного «неудобного» образца искусства Малой Сии и во что обходятся делу погромы, учиняемые в ученой среде амбициозными личностями неустойчивой нравственности.

Презентация предмета искусства. Скульптурное изображение медведя из Малой Сии изготовлено из крупной, правильной, широкой, двухгранной пластины, сколотой с гальки кремненного песчаника (рис. 2). При взгляде на ровную вентральную сторону ее (рис. 2, I) без труда распознается голова животного, оформленная на массивном дистальном конце.

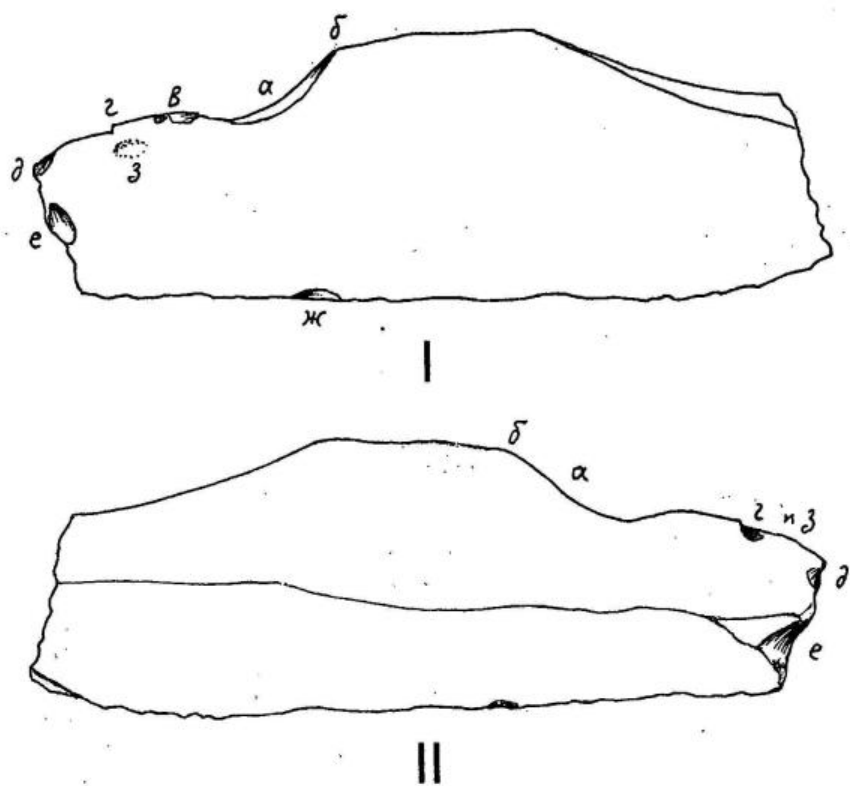


Рис. 2. Скульптура медведя из Малой Сии, изготовленная из двухгранной пластины; I – вентральная сторона; II – дистальная сторона; лунки и прочие знаки не изображены. Здесь и далее рис. В.И. Жалковского.

В частности, *специальными сколами* выделены шея и горб спины, уши, уступ лба, ноздря, открытый рот (возможно, окрашен охрой), а также оконечность передней лапы, которая примыкает к нижнему контуру шеи (см. на рис. 2, соответственно, литеры а, б, в, г, д, е, ж). *Микроточками* выделен вытянуто-овальный глаз (з), расположенный там,

где ему и полагается быть – ниже уступа 2. Те же детали головы и туловища просматриваются и с дорсальной стороны (рис. 2, II).

Знаки числовой информационной системы. Обе поверхности пластины покрыты определенным образом сгруппированными скоплениями знаков – округлой, а также иной конфигурации лунками и микрофасетками краевых сколов (рис. 3, I и II).

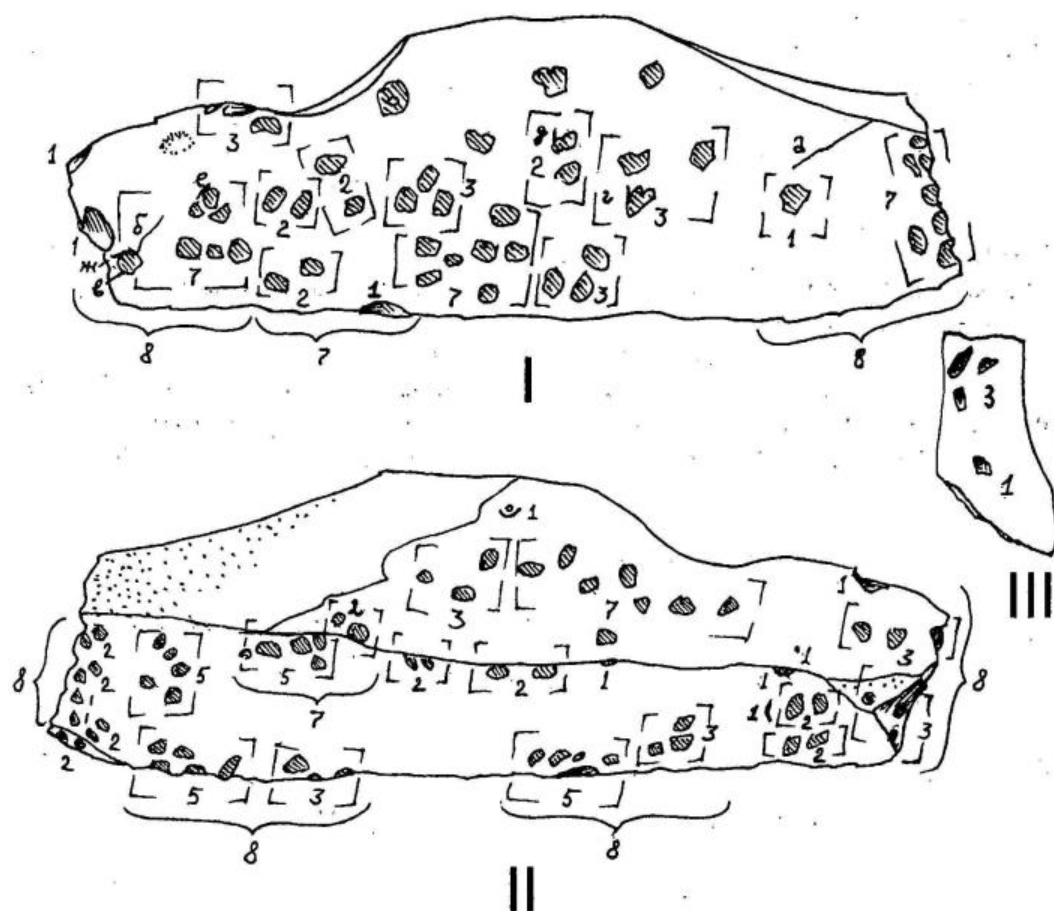


Рис. 3. Группировка знаков на вентральной (I) и дорсальной (II) сторонах пластины; III – знаки на поверхности слома пластины.

Четыре лунки размещены на поверхности слома дистального конца пластины (3 из них расположены близко друг к другу, а 1 – в стороне; см. рис. 3, III). Лунка на широких поверхностях не высверлены, как можно подумать при беглом взгляде на них, а представляют собой каверны, в технологии получения которых еще предстоит разобраться (при консультациях в Геологическом институте АН СССР (Москва) в 1980 г. специалисты высказали гипотезу, что пластина нагревалась, а затем на поверхность ее капали холодной водой; такая идея, однако, сомнительна). Судя по всему, лунки появились при точно рассчитанном, очень сильном и кратковременном (мгновенном!) нажиме на поверхность, в результате чего от пластины отскакивал микроотщеп, массивный в середине и очень тонкий (острый) по периметру (рис. 4). Значительная углубленность каверн в точке мощного нажима подтверждает сказанное. На краях отдельных лунок просматривается, помимо того, тонкие линейные насечки, возможно, сделанные специально до начала операции нажима (см. рис. 3, I, литеры 2, д, е, ж). Кроме того, на той же поверхности выявлены протяженные трещины, оценивать которые трудно (см. там же, литеры а, б, в). Если они искусственного происхождения, то в тексте записи служили, видимо, разделителями, указателями или ограничителями, а если естественного, то, быть может, появились на поверхности как побочное

следствие технологии оформления лунок (сильного, короткого, точечного воздействия на камень). Происхождение красноватой (охристого типа) окрашенности остается неясной (то ли естественная ожелезненность, то ли искусственная окрашенность) и требует специального анализа.

Методические установки. Знаки, совмещенные с предметом искусства, следует воспринимать в качестве определенным образом упорядоченных числовых записей, образующих вместе целостный (единого информационного поля) текст. Содержательная суть его как раз и призвана будет (в случае расшифровки) вскрыть семантический аспект художественного образа. Иначе говоря, *знаковые записи* объекта искусства необходимо воспринимать своего рода *подписью*, разъясняющей идею (концепцию), запечатленную в образе палеолитическим мастером, творцом его. Числа – инструмент познания уникальной мощи и убедительности. Интерпретации, основанные на числовых соображениях, обладают тем, чего напрочь лишены истолкования, основанные на «философического» пошиба мудрствованиях сторонников теории «неудобных фактов».

Одной из главных составляющих предлагаемой методики – определение характера чисел, составляющих текст.

Тестирование числовых знаковых записей на скульптуре медведя из Малой Сии. Лунки на поверхностях пластины объединены в блоки (см. рис. 3). Среди них преобладают группировки по 2, 3, 5, 7 и 8 знаков. Поскольку сходного вида «узоры» («орнаменты») на предметах искусства палеолита обычно представляют собой календарно-астрономические записи, то, приняв за данное, что каждый знак есть символ 1 суток, проведем оценку отдельных блоков, количества знаков на той и другой стороне скульптуры, а затем всех знаков вместе на предмет соответствия их подразделениям лунных месяцев – синодического и сидерического.

Результаты оказались обнадеживающими:

2 знака – период новолуния (невидимости ночного светила, когда оно проходит между Солнцем и Землей);

3 знака – максимальный период невидимости Луны; в течение 3-х суток наблюдается одна и та же фаза ночного светила;

7 знаков – период, кратный с почти идеальной точностью сидерическому (смещение на фоне звезд) обороту Луны : 7 сут. : 27,32 сут. = 0,2562 \approx $\frac{1}{4}$ сид. лун. мес.

и синодическому (смещение Луны относительно Солнца – 7 сут. : 29,5306 сут. = 0,2370 \approx $\frac{1}{4}$ син. лун. мес.*);

8 знаков – период, кратный синодическому обороту Луны : 8 сут. : 29,5306 сут. = 0,2709 сут. \approx $\frac{1}{4}$ син. лун. мес.; в лунных календарях эпохи первобытности обычно использовались чередование 7 сут. \rightarrow 8 сут., когда легкая недостача в первом случае тут же компенсировалась во втором;

49 знаков – количество их на вентральной стороне скульптуры; период этот кратен со значительной точностью синодическому обороту Луны : 49 сут. : 29,5306 сут. = 1,6592 \approx 1 $\frac{2}{3}$ син. лун. мес. (неточность (недостача) – 0,216 сут.); что касается сидерического оборота, то 49 сут. составляют 1 месяц и 22 дня; отмечаю это потому, что именно эти числа позволили начать расшифровку, когда было замечено, что 22 сут. есть число, кратное синодическому обороту Луны (22 сут. : 29,5306 сут. = 0,7449 \approx $\frac{3}{4}$ син. лун. мес.), и это число составляют знаки, расположенные по верхнему краю вентральной стороны скульптуры);

69 знаков – количество их на дорсальной стороне скульптуры; период этот кратен со значительной точностью синодическому обороту Луны – 69 сут. : 29,5306 сут. = 2,3365 \approx 2 $\frac{1}{3}$ син. лун. мес. (неточность (превышение) – компенсация недостачи в 49 сут.!) – 0,2562 сут.); что касается оборота сидерического, то мера кратности всего лишь удовлетворительна (чуть превосходит допустимые 0,02 сут.) – 69 сут. : 27,32 сут. = 2,5256 \approx 2 $\frac{1}{2}$ сид. лун. мес.;

* Палеолитические астрономы работали обычно с точностью \approx 0,02 сут.

118 знаков – общее количество их на обеих поверхностях пластины; период этот с максимально возможной точностью (без учета дробей!) кратен синодическим оборотам Луны: $118 \text{ сут.} : 29,5306 \text{ сут.} = 33,9958 \approx 4 \text{ син. лун. мес.}$ (с точностью 0,367 сут. – недостатка);

122 знака – общее количество их на обеих плоскостях + 4 знака (на поверхности слома пластины); период этот кратен $\frac{1}{3}$ солнечного года – $365,242 \text{ сут.} : 3 = 121,7473 \text{ сут.}$ (превышение – $0,2527 \approx \frac{1}{4} \text{ сут.}$).

Реконструкция порядка считывания блоков и одиночных знаков. Самая сложная часть исследования, изложение которой (с описанием разного рода соображений) невозможно из-за допустимого объема публикации. Напомню лишь, что упомянутое выше подразделение 49 знаков на вентральной стороне пластины на синодические 22 и сидерические 27 позволили в конечном счете установить начальные и конечные точки прочтения текстов на вентральной и дорсальной сторонах, а также направления считывания блоков на каждой из поверхностей (сначала по знакам верхней половины вентрала справа налево, а затем, по нижней его половине, напротив, – слева направо; с дорсальной стороны считывание следует проводить сначала по знакам верхней половины справа налево; далее – по верху ребра – слева направо, а затем по низу его – справа налево; в заключение же знаки нижней половины дорсала считывались слева направо; иначе говоря, считывание велось по зигзагообразным маршрутам, что намекает на синусоидные извивы змеи, древнейшего символа Времени). Результат этой части расшифровки представлен на рис. 5, где стрелки указывают маршруты считывания знаков, а числа 16, 22, 27, 29, 37, 45, 59 – рубежи фазовые и месячные – синодические и сидерические (16 – последние сутки наблюдения полнолуния, канун начала ущерба; $22 \text{ сут.} : 29,5306 \text{ сут.} = 0,7449 \approx \frac{3}{4} \text{ син. лун. мес.}$ – сутки начала наблюдения последней четверти «умирающей» Луны; такой цикл отражает также длительность эпохи лунного затмения, т. е. времени прохождения ночного светила в районе узла его орбиты (пересечения с путем Солнца, когда, случись в это время полнолуние, происходило лунное затмение); 27 – рубеж сидерического лунного месяца; $37 \text{ сут.} : 29,5306 \text{ сут.} = 1,2529 \approx 1 \frac{1}{4} \text{ син. лун. мес.}$; $45 \text{ сут.} : 29,5306 \text{ сут.} = 1,5238 \approx 1 \frac{1}{2} \text{ син. лун. мес.}$; $59 \text{ сут.} : 29,5306 \text{ сут.} = 1,9979 \approx 2 \text{ син. лун. мес.}$). Этот «унылый» каскад цифири призван обратить внимание на два важных обстоятельства:

1 – календарист Малой Сын выводил на заметные позиции знаки, определяющие рубежи месяцев, а также примечательных частей их;

2 – считывание лунного времени велось с учетом синодических (фазовых) и сидерических (звездных) вариантов его (как в случае с ачинским жезлом [Ларичев, 1987 б], что можно объяснить лишь стремлением предсказать (или, может быть, вычислить?) момент, когда может наступить затмение Луны или Солнца). Что так оно и было подтверждают реконструкции разных по виду годовых систем счисления времени.

Реконструкция счисления лунного года. Трехкратный проход по знакам 49 и 69 выводит на рубеж окончания лунного года –

$$[49 + 69 \text{ сут.}] \times 3 = 354 \approx 354,367 \text{ сут.}$$

Реконструкция счисления солнечного года, когда за основу принимается лунный счет времени. После завершения счисления трех лунных лет по установленной схеме, в счетную систему вводился интеркалярный – 34 дня (завершение этого периода определяет заметный знак – (см. рис. 5, I, 34). Это операция и позволяла выравнивать лунный счет времени с времени солнечным. В самом деле,

$$[354 \text{ сут.} \times 3] + 34 \text{ сут.} = 1096 \text{ сут.}$$

$$1096 \text{ сут.} : 365,242 \text{ сут.} = 3,000750 \text{ солн. лет.}$$

Цикл в 34 дня есть синодический период от последнего дня наблюдения серпа «умирающей» Луны на Востоке, до неомении – первого явления на Западе серпа «народившейся» Луны. Этот же период интеркаляции отражает продолжительность эпохи солнечного затмения, когда при прохождении невидимой Луны (новолуние!) в районе узла ее орбиты может случиться одно или даже два затмения дневного светила.

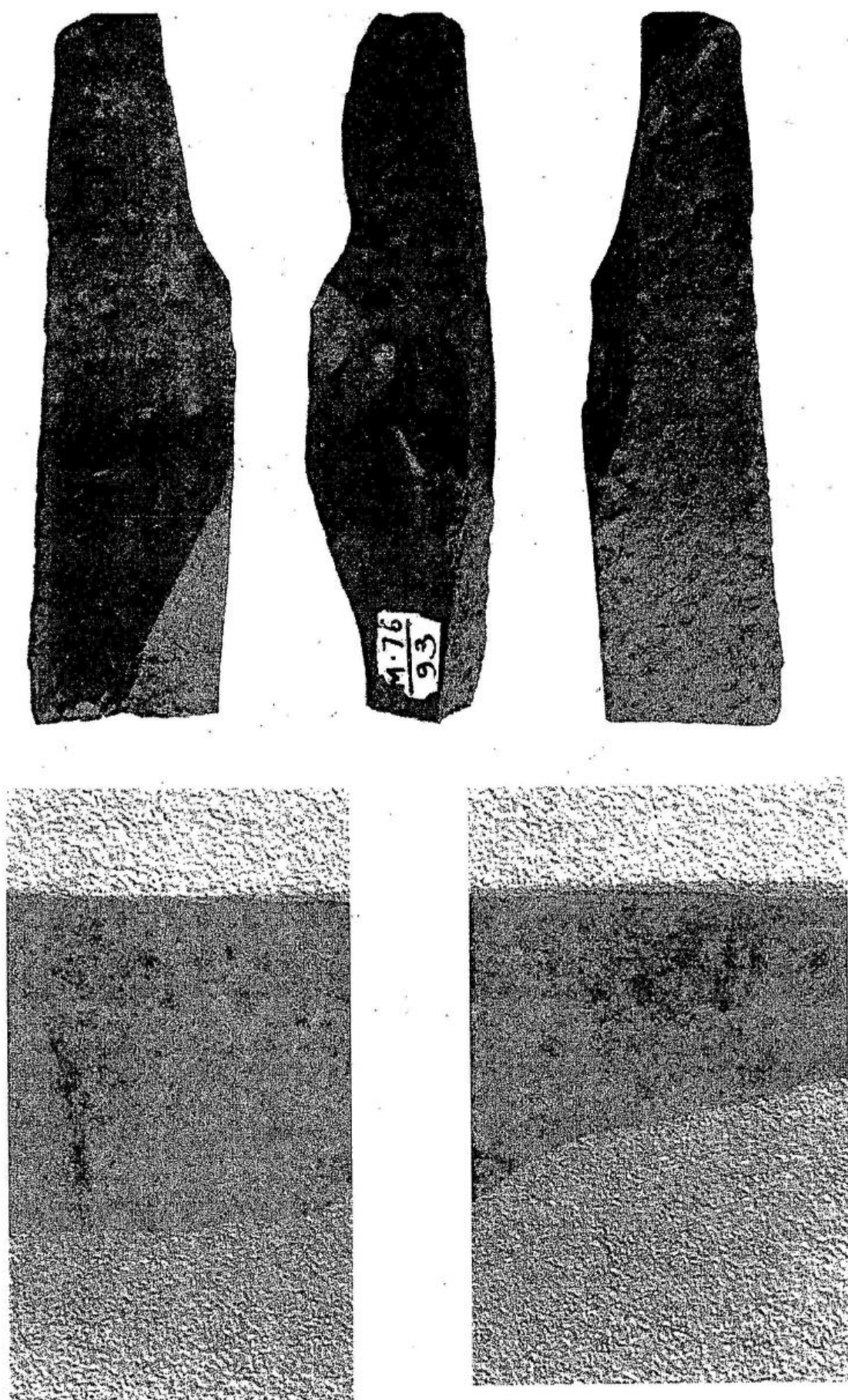


Рис. 4. Вентральная и дорсальная стороны пластины. Фото В.Н. Кавелина.

Это сказано с целью дополнительного подтверждения идеи, объясняющей причину отсуживания Луны не только относительно Солнца, но и звезд.

Реконструкция счисления солнечного года, когда за основу принимается цикл, продолжительность которого составляет 122 дня $\approx \frac{1}{3}$ тропического года (118 знаков на поверхностях пластины + 4 знака на поверхности слома). Трехкратный проход по всем знакам цикла 122 выводил на рубеж солнечного года —

$$122 \text{ сут.} \times 3 = 366 \text{ сут.} \approx 365,242 \text{ сут.}$$

Выравнивание превышения производилось, видимо, в точности так же, как в солнечных календарях Мальты: после считывания трех лет продолжительностью 366 сут., 4-й год принимался равным 363 дня, что осуществлялось посредством неучета 3-х знаков на поверхности слома (1 \rightarrow 3!). Палеолитический календарист Малой Сии следовал истине следующих формул

$$[366 \text{ сут.} \times 3] = 1098 \text{ сут.}$$

$$1098 \text{ сут.} + 363 \text{ сут. (укороченный год!)} = 1461 \text{ сут.}$$

$$1461 \text{ сут.} : 365,242 \text{ сут.} = 4,00008 \text{ солн. лет.}$$

Иные годовые циклы, сезон начала счисления времени и биоритмы. У меня нет полной уверенности, что малосийский астроном специально отмечал иные годовые циклы. И все же обращаю внимание на позиции лунок 328, 173 и 347, определяющих, соответственно, рубежи звездного (сидерического), половины и целого драконического (затменного) годов. Мне кажется, что позиции эти — заметны; что касается биоритмов, то заслуживают внимания два обстоятельства:

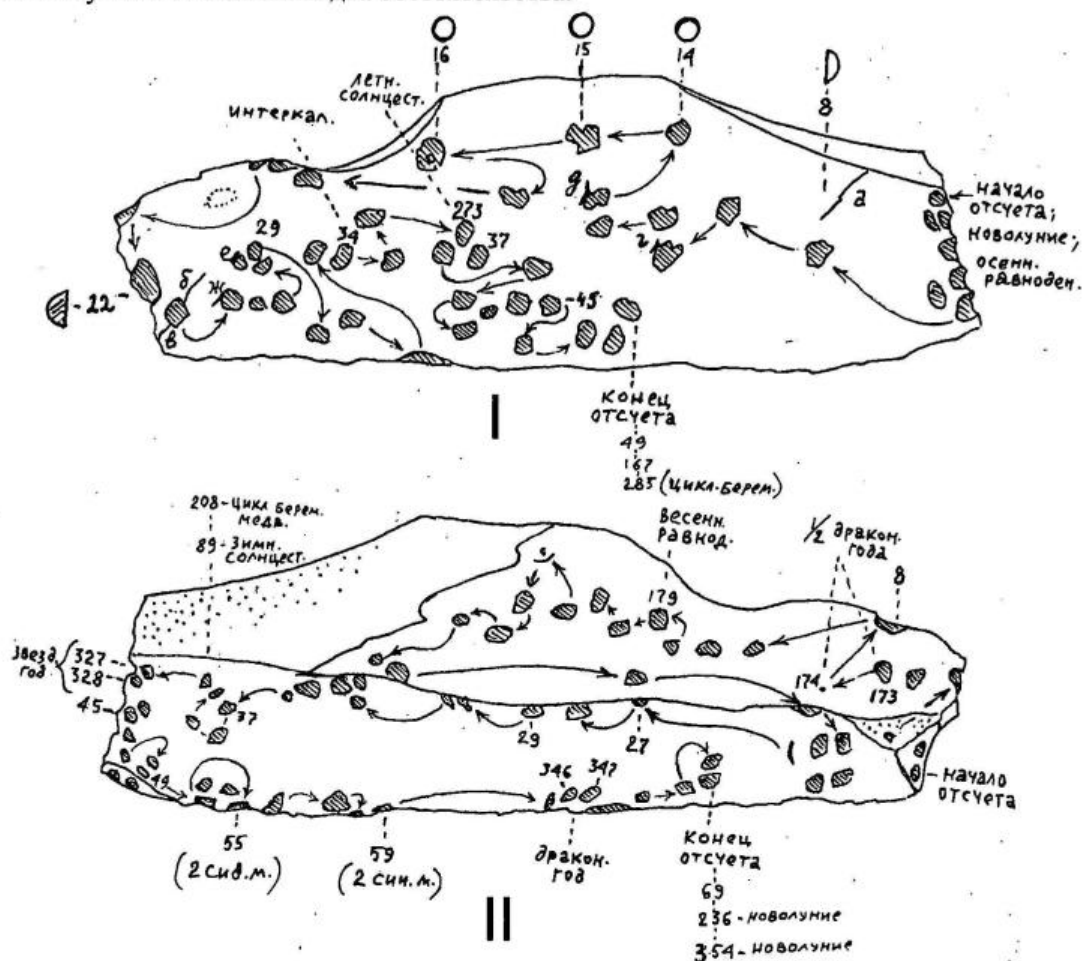


Рис. 4. Маршруты счисления знаков. Цифры обозначают рубежи разного рода циклов.

1 – на последний знак вентральной поверхности пластины приходится не только 49-й день, но и (при повторных проходах) – 285-й, что есть *цикл беременности женщины*

285 сут. : 29,5306 сут. = 9,6510 \approx 9 $\frac{1}{3}$ *син. лун. мес.*;

2 – на заметный знак приходится при первом проходе 89-е сутки, что есть период от осеннего равноденствия до зимнего солнцестояния (см. на рис. 5 цифру 89, которая определяет последний знак блока 5); это означает, что *новогодие у малосыйцев определяло осеннее равноденствие* (другие рубежные (сезонные) знаки – 179 (весеннее равноденствие), 273 (летнее солнцестояние)); 3 – на тот же знак 89 при повторном проходе приходятся 208-е сутки от начала счисления, что есть цикл беременности медведя (*то же самое (с разностью в 1 сутки)* зафиксировано в счетной системе скульптуры медведя из Новой Сыды, которую, кстати, составляют 69 знаков (см. рис. 1, з), число, в точности соответствующее количеству знаков на дорсальной стороне пластины из Малой Сыи!).

Семантика образа медведя в искусстве Малой Сыи. Значимость животного представляется в основе той же, что представлено в интерпретациях медведя в искусстве культур палеолита Западной Европы [см. Ларичев, 2002] и в скульптурном образце Новой Сыды [Ларичев, Аннинский, 2004; в печати]. Отсылая читателя за подробностями к этим публикациям, здесь отмечу лишь главное. Зверь этот – не земной, а небесный, божественного, жертвенного (во имя творения Мира?) и дионисийско-осирического, по сути своей, ранга (олицетворение нескончаемого круговорота жизни, смерти и воскрешения). Он же и носитель сакральных знаний – естественно-научных, культово-обрядовых и религиозных. Он и, вследствие «календарности» и «астрономичности» своей, видится божественным регулятором хозяйственно-экономических действ. Медведь Малой Сыи представлен в жертвенной позе – лежащим, с подогнутыми под тело лапами, с открытым ртом, который символизирует знак четверти «умирающей» Луны» (и число (22-й день!), и контур фасетки намекают на то). Открытый рот и четко обозначенная ноздря, символизирующие дыхание (жизнь), направлены, видимо, в сторону нечто такого, что обеспечивает возрождение умершего (так изображались головы медведей в искусстве палеолита Западной Европы (см. на рис. 1, ж–л [Ларичев, 2002]). Как можно убедиться, ноздри и рот медведей устремлены в сторону символов «подателей жизни» – фаллоса (рис. 1, к), мирового древа (рис. 1, з), «травы жизни» или «живой воды» (рис. 1, ж, и, л).

Краткие итоги. Скульптурное изображение медведя из Малой Сыи – первоклассный источниковый материал для реконструкций самого сложного аспекта жизни палеолитического человека – его интеллектуального и духовного мира. Еще две скульптуры Малой Сыи с такого же типа знаковыми записями календарно-астрономического содержания подготовлены к публикации. Они позволят продолжить изучение этих аспектов. Теперь же можно со всей определенностью констатировать верность идеи высокого уровня астрономических знаний и арифметики в культурах раннего этапа верхнего палеолита Сибири. Они, эти знания, как бесценное наследие десятков тысячелетий работы ума предка над разгадками тайн Природы, запечатлевались посредством *своеобразной письменности ледниковой эпохи* – в образах и знаках того, что современный археолог воспринимает по привычке искусством.

Литература

Авдеев С.М. Толбагинская скульптура как образец первобытного искусства. Четвертичная геология и первобытная археология Южной Сибири. Тезисы докладов. Всесоюзная конференция (2-8 июня 1986 г., Улан-Удэ). Часть II. Улан-Удэ, 1986.

Герасимов М.М. Мальта – палеолитическая стоянка. Иркутск, 1931а.

Герасимов М.М. Палеолитическая стоянка в Мальте // Сообщения Государственной академии истории материальной культуры. – 1931, № 11-12.

Грязнов М.П., Столяр А.Д., Рогачев А.Н. Письмо в редакцию // Советская археология, 1981, № 4.

- Константинов М.В. и др.** Скульптуре медведя из Забайкалья 30 тысяч лет // Природа, 1981, № 11.
- Константинов М.В. и др.** Древнейшая скульптура Сибири // Краткие сообщения Института археологии, 1983, № 173.
- Ларичев В.Е.** У истоков верхнепалеолитических культур и искусства Сибири (к открытию в Кузнецком Алатау поселения Малая Сыя и скульптурного изображения черепахи // Рериховские чтения. 1976 год. – Новосибирск, 1976.
- Ларичев В.Е.** Скульптура черепахи с поселения Малая Сыя и проблема космогонических представлений верхнепалеолитического человека (описание находки и опыт предварительной интерпретации) // У истоков творчества. Первобытное искусство. – Новосибирск, 1978а.
- Ларичев В.Е.** Искусство верхнепалеолитического поселения Малая Сыя: датировка, виды его и образы, их художественный стиль и проблема интерпретации (предварительное сообщение) // Известия СО АН СССР. Серия общественных наук. – 1978б, № 11, вып. 3.
- Ларичев В.Е.** Мамонт в искусстве поселения Малая Сыя и опыт реконструкции представлений верхнепалеолитического человека Сибири о возникновении Вселенной // Звери в камне. Первобытное искусство. – Новосибирск, 1980а.
- Ларичев В.Е.** Зооантропоморфная скульптура рожающего существа верхнепалеолитического поселения Малая Сыя и Великая Богиня – Мать индийской мифологии // Рериховские чтения. 1979 год. Новосибирск, 1980б.
- Ларичев В.Е.** Скульптурное изображение женщины и лунно-солнечный календарь поселения Малая Сыя (семантика образа и реконструкция способа счисления времени на раннем этапе верхнего палеолита Сибири // Известия СО АН СССР. Серия: история, филология и философия. – 1984, № 3, вып. 1.
- Ларичев В.Е.** Двуконечная зооантропоморфная скульптура из камня поселения Малая Сыя (явление полиобразности в мобильном искусстве раннего этапа верхнего палеолита Сибири и проблема семантики первобытного художественного творчества) // Древности Сибири и Дальнего Востока. – Новосибирск, 1987а.
- Ларичев В.Е.** Ачинский жезл и его знаковая система // Исследования памятников древних культур Сибири и Дальнего Востока. – Новосибирск, 1987б.
- Ларичев В.Е.** Мудрость Змеи: Первобытный человек, Луна и Солнце. – Новосибирск, 1989.
- Ларичев В.Е.** «Белая Лошадь» – святилище древнекаменного века Хакасии (астрономические аспекты памятника и астральная подоснова искусства древнекаменного века Сибири). – Новосибирск, 1992 (см. стр. 78 и рис. 26–28).
- Ларичев В.Е.** Лунные и солнечные календари древнекаменного века // Календарь в культуре народов мира. – М., 1993.
- Ларичев В.Е.** Большое видится на расстоянии (документы истории открытия и первых лет раскопок Мальты) // Вестник антропологии. Альманах. – 1998, вып. 5.
- Ларичев В.Е.** Звездные боги: Слово о великих художниках – созерцателях Неба, мудрецах и кудесниках. – Новосибирск, 1999.
- Ларичев В.Е.** Семантика образа медведя в искусстве древнекаменного века (календарно-астрономический и космогонический аспекты) // Традиционная культура Востока Азии. Выпуск четвертый. – Благовещенск, 2002.
- Ларичев В.Е.** Знаки и образы искусства палеолита в семантическом русле течения времени (прочтение числовых текстов на «bâton de commandement» из Ла Марша) // Религиоведение, 2003, № 2.
- Ларичев В.Е., Аннинский Е.С.** Скульптуры медведя и мамонта раннего этапа верхнего палеолита Енисея (полиобразность предметов искусства малых форм древнекаменного века Средней Сибири и семантика их). В печати.
- Любин В.П., Абрамова З.А., Борисковский П.И.** Приложение. Результаты ознакомления специалистов с материалами В.Е.Ларичева // Советская археология, 1981, № 4.
- Матющенко В.И.** Триста лет истории сибирской археологии. Т. II. Омск, 2001.

Окладников А.П. За методологию диалектического материализма в истории доклассового общества // Сообщения Государственной академии истории материальной культуры. – 1932, № 3–4.

Петри Б.Э. Сибирский палеолит. – Иркутск, 1923.

Петри В.Э. Сибирский палеолит. Атлас. Рисунки выполнены М.М. Герасимовым. – Вып. 1. Ангарский палеолит. – Иркутск, 1927.

Петри Б.Э. Далекое прошлое Прибайкалья. – Иркутск, 1928.

Научное издание

Информационные технологии в гуманитарных исследованиях

Выпуск 6

Ответственный редактор:
академик РАН, д.и.н. *Ю.П.Холюшкин*

Компьютерная вёрстка – *В.Т.Воронин*

Подписано в печать 03.12.2003 г.
Формат 60×84 1/8. Офсетная печать.
Уч.-изд. л. 12,5. Тираж 200 экз.
Заказ № 582

Лицензия ЛР № 021285 от 6 мая 1998 г.
Редакционно-издательский центр НГУ.
630090, Новосибирск-90, ул. Пирогова, 2.

